

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 13:31:44 Search time 3097 Seconds  
(without alignments) 10233.437 Million cell updates/sec

Title: US-08-468-190C-1\_COPY\_9874\_10962

Perfect score: 1089  
Sequence: 1 ATGGCAGCGCTACTTGAGAC.....CCGACCTCATATAAATGCTGA 1089

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database:

GenBank: 1: gb\_ba:\*  
2: gb\_btg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_iny:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_man:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htg\_hum:\*  
40: em\_htg\_mus:\*  
41: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID          | Description          |
|------------|--------|-------------|--------|----------------|----------------------|
| 1          | 1087.4 | 99.9        | 18912  | 14 ILU28832    | U28832 Infectious    |
| 2          | 63     | 5.8         | 1627   | 14 HSMGDELCO   | L31865 Gallid hepi   |
| 3          | 51.2   | 4.7         | 125020 | 9 AF429315     | AF429315 Homo sapi   |
| 4          | 43.2   | 4.0         | 1246   | 6 AX164174     | AX164174 Sequence    |
| 5          | 41.4   | 3.8         | 138025 | 6 AP003261     | AP003261 Oryza sat   |
| 6          | 41.4   | 3.8         | 200183 | 8 AP003227     | AP003227 Oryza sat   |
| 7          | 40.2   | 3.7         | 125020 | 9 AF429315     | AF429315 Homo sapi   |
| 8          | 39.6   | 3.6         | 12977  | 2 DMU19099     | DMU19099 Drosophila  |
| 9          | 39     | 3.6         | 110000 | 2 LMFICHR18_04 | Continuation (5 of   |
| 10         | 38.8   | 3.6         | 117004 | 2 AP003743     | AP003743 Oryza sat   |
| 11         | 38.8   | 3.6         | 130732 | 2 AP003848     | AP003848 Oryza sat   |
| 12         | 38.2   | 3.5         | 16984  | 1 AE006954     | AE006954 Mycobacte   |
| 13         | 38.2   | 3.5         | 40051  | 1 MSGY224      | AD000004 Mycobacte   |
| 14         | 38.2   | 3.5         | 40838  | 1 MY25D10      | Z95558 Mycobacteri   |
| 15         | 38.2   | 3.5         | 109554 | 2 CNO80C90     | AL732649 Oryza sat   |
| 16         | 38     | 3.5         | 4543   | 3 DROCSM       | M94730 Drosophila    |
| 17         | 38     | 3.5         | 38889  | 3 DBRN25G24    | AL132797 Drosophila  |
| 18         | 38     | 3.5         | 45760  | 2 AC017610     | AC017610 Drosophila  |
| 19         | 38     | 3.5         | 177724 | 3 AC104146     | AC104146 Drosophila  |
| 20         | 38     | 3.5         | 291976 | 6 AX251969     | AX251969 Sequence    |
| 21         | 37.6   | 3.5         | 6310   | 6 AX344361     | AX344361 Sequence    |
| 22         | 37.6   | 3.5         | 6310   | 6 AX348754     | AX348754 Sequence    |
| 23         | 37.6   | 3.4         | 1172   | 2 LMFICHR32_20 | L26252 Leishmania    |
| 24         | 37.4   | 3.4         | 110000 | 2 LMFICHR32_21 | Continuation (21 o   |
| 25         | 37.4   | 3.4         | 110000 | 2 LMFICHR32_21 | Continuation (21 o   |
| 26         | 37.4   | 3.4         | 110000 | 2 LMFICHR32_21 | Continuation (21 o   |
| 27         | 37.2   | 3.4         | 10177  | 1 ABO04977     | ABO04977 Halobacte   |
| 28         | 37     | 3.4         | 10261  | 1 ABO011997    | ABO011997 Xanthomon  |
| 29         | 37     | 3.4         | 234840 | 1 SMO5012      | AL591789 Sinothizo   |
| 30         | 37     | 3.4         | 345847 | 1 APO03012     | AP003012 Mesothizo   |
| 31         | 36.8   | 3.4         | 30000  | 1 ABO05904     | ABO05904 Arabidoste  |
| 32         | 36.6   | 3.4         | 6176   | 3 AF104400     | AF104400 Drosophila  |
| 33         | 36.6   | 3.4         | 12177  | 3 AF104401     | AF104401 Drosophila  |
| 34         | 36.6   | 3.4         | 71516  | 2 AC014220     | AC014220 Drosophila  |
| 35         | 36.6   | 3.4         | 155263 | 8 AP003224     | AP003224 Oryza sat   |
| 36         | 36.6   | 3.4         | 160001 | 3 AC113619     | AC113619 Drosophila  |
| 37         | 36.6   | 3.4         | 189976 | 3 AC023710     | AC023710 Drosophila  |
| 38         | 36.6   | 3.4         | 305018 | 3 AE003437     | AE003437 Drosophila  |
| 39         | 36.2   | 3.3         | 40476  | 1 SCC75A       | AL133220 Streptomy   |
| 40         | 36     | 3.3         | 10732  | 6 E32986       | AC106407 Gene encodi |
| 41         | 36     | 3.3         | 161205 | 2 AC106407     | AC106407 Rattus no   |
| 42         | 35.8   | 3.3         | 58270  | 9 AF250325     | AF250325 Homo sapi   |
| 43         | 35.8   | 3.3         | 115666 | 2 AC105744     | AC105744 Oryza sat   |
| 44         | 35.8   | 3.3         | 153472 | 9 ABO20868     | ABO20868 Homo sapi   |
| 45         | 35.8   | 3.3         | 176967 | 2 AC022742     | AC022742 Homo sapi   |

## ALIGNMENTS

| RESULT 1   | LOCUS    | DEFINITION                          | ACCESSION | VERSION | KEYWORDS   | SOURCE | ORGANISM                            | REFERENCE | AUTHORS                              |
|------------|----------|-------------------------------------|-----------|---------|------------|--------|-------------------------------------|-----------|--------------------------------------|
| ILU28832   | ILU28832 | 18912 bp DNA linear VRL 09-AUG-1996 | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| LOCUS      | ILU28832 | 18912 bp DNA linear VRL 09-AUG-1996 | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| DEFINITION | ILU28832 | 18912 bp DNA linear VRL 09-AUG-1996 | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| ACCESSION  | ILU28832 | 18912 bp DNA linear VRL 09-AUG-1996 | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| VERSION    | U28832   | U28832                              | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| KEYWORDS   | U28832   | U28832                              | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| SOURCE     | U28832   | U28832                              | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| ORGANISM   | U28832   | U28832                              | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| REFERENCE  | U28832   | U28832                              | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |
| AUTHORS    | U28832   | U28832                              | U28832    | U28832  | GI:1486484 | U28832 | Infectious laryngotracheitis virus. | 1         | Wild, M.A., Cook, S. and Cochran, M. |

|               |   |
|---------------|---|
| TITLE         | A genomic map of infectious Jaryngotracheitis virus and the sequence and organization of genes present in the unique short and flanking regions |
| JOURNAL       | Virus Genes 12 (2), 107-116 (1996)  |
| MEDLINE       | 97033380  |
| PUBMED        | 8879127   |
| REFERENCE     | 2 (bases 1 to 18912)  |
| AUTHORS       | Wild, M.A.  |
| JOURNAL       | Direct Submission   |
| TITLE         | Submitted (09-JUN-1995) Martha A. Wild, Syntro Research Laboratory,   |
| FEATURES      | 3535 General Atomics Court, San Diego, CA 92121, USA  |
| SOURCE        | Location/Qualifiers   |
| repeat_region | 1..18912  |
| repeat_region | /organism="Gallid herpesvirus 1"  |
| repeat_region | /strain="USDA challenge strain"   |
| repeat_region | /db_xref="taxon:10386"  |
| repeat_region | <1..2909  |
| repeat_region | /note="large inverted repeat flanking the short unique region"  |
| repeat_region | /rpt_type="inverted   |
| repeat_region | 697..1533   |
| CDS           | /codon_start=1  |
| CDS           | /product="US10"   |
| CDS           | /protein_id="AAC55093.1"  |
| CDS           | /db_xref="GI:1486485"   |
| CDS           | /translation="MENMLDGCYPLALMDSPTTAHAAPNGERRRQGAASASADS  |
| CDS           | VDPCIRIASRLRELISELKDGEFTSARDNRRLIANERLSAFIGASRATR   |
| CDS           | GLGRLPMASVESVANSPTDNGNGELGELEAMEGTEDFMLDSDGREDSSRM  |
| CDS           | OSEMREVIKEELISLMSRLPALASERTSRSLAAGMCCLMHPRERECATD   |
| CDS           | SIYVSLFCVGTGRVPOSEMRREYLAALRAGAANAASPEVSASIFARDGIALALAL   |
| CDS           | RR"   |
| CDS           | complement(1900..2784)  |
| CDS           | /note="coding region unique to avian herpesviruses"   |
| CDS           | /codon_start=1  |
| CDS           | /product="unknown"  |
| CDS           | /protein_id="AAC55094.1"  |
| CDS           | /db_xref="GI:1486486"   |
| CDS           | /translation="MSKCYCLARHLKSPRCVGRRAVFGGLATNSRPTSHLDAFSA   |
| CDS           | AFRTDLPGRGFMRASQSCDIEFWPDIAAVVQARAVFEGERLGSLOVEDITAD  |
| CDS           | PRIPAAKRAVAALVGLMTALSELVNGELSEKVGKQIPRAAWEIRDPVPI   |
| CDS           | GPDLISFEESAVALPELVYIRAGGASHRSAMNNOSSAPAGLAIRIGMVRSLIV   |
| CDS           | IAPLSNFTLPEDLPBGSONSIRAFVAMNCVATDKIMSDVAVPVEBSYSHCLAE   |
| CDS           | IIMERAFCTCPNPPPKM"  |
| CDS           | 2910..16003   |
| CDS           | /note="short unique region"   |
| CDS           | complement(2916..3605)  |
| CDS           | /codon_start=1  |
| CDS           | /product="US2"  |
| CDS           | /protein_id="AAC55095.1"  |
| CDS           | /db_xref="GI:1486487"   |
| CDS           | /translation="MAPVKYIVASAVDSHYKLPNSRELSDSGMKELVHAVKMASYD  |
| CDS           | RPSYLVIVRASLYEVSGELSLPRMCRPIVIRGEGGDPGVSPEMSGLDAGYHS  |
| CDS           | SGAYAAKEFLHMLVGTADICMALNLPAPKTELTTEGKRGVEVFLVNGKRTL   |
| CDS           | SLSHPSVMTLAPSSLRTPMPYIVKFLKYPNAACVSDSDGENDQKPTFLSLRK  |
| CDS           | SKKRPSPRRR"   |
| CDS           | 3694..5124  |
| CDS           | /codon_start=1  |
| CDS           | /product="protein kinase"   |
| CDS           | /protein_id="AAC55096.1"  |
| CDS           | /db_xref="GI:1486488"   |
| CDS           | /translation="MFRRLICSRRAKRRRTTENPLTSKRCVYLDSPSRMSLRPYA   |
| CDS           | EILPTAGVEKRLAVSVLTERAEPTENTAVNSIPPANENGQNFAYADGSPSTTK   |
| CDS           | VDGSHTFDEASDYAGVPLAQTRLRHSDSEFLQHFVLDLVEAGVGLCDVARYTE   |
| CDS           | EEORRRGVNSTNGSKSKRLIAKYVNGSTRAASQLENEILVGRLNHEVLIQOELL  |
| CDS           | RYPPNTYMLTOROFLDYSYMDAEAFMKDSPMLKORIRIMKOLMASVSYIHSKKLIH  |
| CDS           | BDIKLENTIFLNCCKTVLGDGCVTTPRENGRPREYMGVGTAAVNSPEILLAROSYCI   |
| CDS           | TDIIRSCGVVLEMYSHFECPIGOGGGPHOOLKLVIOISLSCDEEFPDNLNVIYH  |
| CDS           | YASIDRAGHTVPSLIRNLHLPADEYPLVYMLTFDMRLRPSAAVLAAMPLFSAEERT  |
| CDS           | ITIHGKHKPIREIRARVPRSMSEG"   |
| CDS           | 5210..7081  |
| CDS           | /codon_start=1  |
| CDS           | /product="uL47"   |
| CDS           | /protein_id="AAC55097.1"  |
| CDS           | /db_xref="GI:1486489"   |
| CDS           | /translation="MTPLRLTKRPPARRCVSFVIHYSETKLDTRYNKTMLLYRPDST   |
| CDS           | MHSGGANHNGIRPRRSIGAFSAREKTGRNRLSSSSSDMLDPEFSDKEGKGM   |
| CDS           | TYDDPADITAEVLSOANDVLOLYHEDPAEPRRYVRESKPTPIQFNAMPQSPNAD  |
| CDS           | FTRAPYTPPSAEVVDALRLKAGAVSKVYLQCVPTPRSKSKSTTVYDADSPFSDSWRLH  |
| CDS           | ONVTVVLRQHTLRRLHPSKSYVNCNDPLVAVANOFHMGKMPVSLKASPEENI  |
| CDS           | OHGMAAVFRNAGGLSEFLWPMARAAFEERKRLRLACLSDLDIDAVALASFPYWMG   |
| CDS           | VODTSREFPALGCLSEFVALVLAETVLAETVLAETVLAETVLAETVLAETVLAETVLA  |
| CDS           | VKNESYNGAGCTLRGIVASNTATVAVCANVYSTIRKSDVATACVAVRTETLA  |
| CDS           | RRPMSDIYRIQKEFFFYIAMORVANTNFCNIIKRSIDVACAPFLIPASSEKRL   |
| CDS           | QOLKMLCPILVPIQYEDFSKMSSELKRELETFVKAISDRPDRSLRLISDHR   |
| CDS           | EILADVFRKPVIDEVRAVALSTAAAGVAKARLTVRAVPAGAVARRKSEI"  |
| CDS           | 7245..8123  |
| CDS           | /codon_start=1  |
| CDS           | /product="glycoprotein G"   |
| CDS           | /protein_id="AAC55098.1"  |
| CDS           | /db_xref="GI:1486490"   |
| CDS           | /translation="MSGFSNIGSIATVSLVCSLLCASVIGAPVLDLESSPPFGKI   |
| CDS           | IAQACNRTTTEVTPWMSDYSGRTEGVSEVAKWFNGSNPESFVGVSETEGSHEDLS   |
| CDS           | TCMALINHLNLSVCRASDAGIPDEKQCEKQRLRSGVEIGSYSGNSGLVLYPGMY  |
| CDS           | DAGIYVQLSVGKGYGSGYLDVGPBGCHDOYGYTYSLADEASPLSDVAPSL  |
| CDS           | DGMBEDYSNCLDMPRLRPPTYVCSHVEDENATDILYLMDACAPLDDEYDERSE   |
| CDS           | TPRMVVFSPSTLQ"  |
| CDS           | 8333..11290   |
| CDS           | /note="ORF5"  |
| CDS           | /codon_start=1  |
| CDS           | /product="AAC55099.1"   |
| CDS           | /protein_id="AAC55100.1"  |
| CDS           | /db_xref="GI:1486491"   |
| CDS           | /translation="MGTMVLRLLEFLAVADAALPTGFCRWKYPPEGCTIOENLVIA  |
| CDS           | ESPVGTGATYPPPEGAVSFQIFADPTPLRIRYGATEDELELERTSASDADNVLVLS  |
| CDS           | YRPRBEIHGAFTIGVFATGOSTESSYSRVLVNALSLEVSALTEPCDENFQONP   |
| CDS           | TWGSKRWLGPSPYVRNDVAVLKAYIGIECYNSAAGCTGLNLTNTPFYSPVPA  |
| CDS           | VTTGTGSPSRITVYSRENGOVLNNSGFLVKTPTDIDGRAMVIANVANSPADS  |
| CDS           | GSVLAFTAPRGRKLPASAIOLHIDMSCTEPPCTTDCQKMETPYRALGSPARD  |
| CDS           | SIRGATLPEDPTAAPPDGTSTPTPTVYBPATTTILPSTDSMGFFSTARAGST  |
| CDS           | LSVPOETDRITLPTPLPLTTPRESNTLPTAPISLTPESAHEHTQSOAEV   |
| CDS           | LVFQSPSTESERTLASQSEWYFTQSTEQALTOQILETALFTQTSABQMETT   |
| CDS           | OTPAETEAAPQAPSTIPETIQSGESTPPTERASAPAEVFTQSSSTVEVFPQTS   |
| CDS           | TVPTLTSSSTSEPIAIFRTQSAGAEATQTSASAPDAPROSTHTFFQTSVSPVA   |
| CDS           | TORPSTESLQSTPEPPTPTLGAPEIQTQTPAAAEVYRSSSTPETERQST   |
| CDS           | LASQNPSTSGQNTPEPPTPTPTQTKVTEKNTLSPPVYSEFHSMTAESQT   |
| CDS           | PLDVKVYKFSNDGEVYATCVSTVSPYRVEENKAVDLYVDWDEISGSPACVFS  |
| CDS           | NEWKQOLYTRVDGRTSVQMLCSYSHSPEPCLDTSLIAEKDIABEFLPTSDP   |
| CDS           | OTAYCTTLPSPGVPRFWSLNVSLPEVLTATVVSHTAGOSTVMKSSARAGAWIS   |
| CDS           | GRGNIEYCTVLISDGTTRTKERCCLTNMIAVENGAQAOLYSLFSGLSGCSI   |
| CDS           | SALYATLMTAIF"   |
| CDS           | 11098..12402  |
| CDS           | /codon_start=1  |
| CDS           | /product="glycoprotein D"   |
| CDS           | /protein_id="AAC55101.1"  |
| CDS           | /db_xref="GI:1486492"   |
| CDS           | /translation="MHRPHRRHSRYAAKEVNLKHMDCGKRCGGAAYFTLEMTVC  |
| CDS           | RIMHEICFVRNMDRHLFLRNAMETVILSSPASOSTAATVYDYLIGRALDALATIP   |
| CDS           | ANGRYNNYLRVSGCYVELNPLSNVDDMIISAKKEKKGCPREASVYVYIKGDG  |
| CDS           | EDKTCPIYKREYRECGVQLSECAVDSAAVMADVPLSVLSRNGAGLTIFSPALV   |
| CDS           | SGOYLTLTKIGRQAQTLVTLVENDRLKIGSOLNPLPSKCTWDETOYTFQGEHLV  |
| CDS           | IADTNTRHADVYRGYEDILQRMNNILRRKNPSAPDPSVPOEIPAVYAKKAGRP   |
| CDS           | DAESSEKAPPEDESDMOAEGENPAALPEDEVEDPEEDHNDPDYVNDMPAV  |
| CDS           | IPVEETKSSNAWMPIFAAVACAVALLVGLWVSIYKARS"   |
| CDS           | 12510..13598  |
| CDS           | /codon_start=1  |
| CDS           | /product="glycoprotein I"   |
| CDS           | /protein_id="AAC55101.1"  |
| CDS           | /db_xref="GI:1486493"   |
| CDS           | /translation="MSLSGLTALLAATLAPGAMGIVTGNHVSARIDDDHIVYAP  |
| CDS           | RPEATIDQLFPMGQRPKRPYSGYVAVARSDITNOCYBELSEERENTHSSSVF  |
| CDS           | VGVKVEITYTSASRLGLGPPHFKLTIIRNRPNDGSEFYIVRLDDKEITDVAIOL  |
| CDS           | SVQOFANTATRGRLYSKASCRTGLPVLQLEALVLTREESNRQAVYATATTSAA   |
| CDS           | TPTPTVATIGASELEAHEFTFPMLENDVDIYEPTPAENSNVTYRLGTMSPTLIGYV  |
| CDS           | AAVYSATIGLIVISIVTRMNCVPHRKLVDVQDDDEERSQVTRRESKFGMVACEIK   |
| CDS           | GADODSELVELVAIVNSALSPDSIKM"   |
| CDS           | 13792..15291  |
| CDS           | /codon_start=1  |

```
/product="glycoprotein E"  
/protein_id="AAC5102.1"  
/db_xref="GI:1486494"  
translation="MNLVIVLASCILARLTETRVLELSTQAVLGEEDPRNVEGT  
VIRKTRVIRNACMKKADVCSSPNCFHDLIDYGGKRCPPAPLSANVILKNGES  
FVLGSGHNSININIMTEYGLLPEDTRSEDEIYFRRIISODPLAMETSYNL  
SHVDEKAPAEVEIDTIRKSEAHAVELQMLPHELDNSPTVPLRVPRVGRVPS  
KENVYMWGFDVKEECVEYKLEPCVYPTGKQOPATNQLIGSLVMEFLGAA  
SLIDCSRDLECHENRVPRLRFDSLSERAGLISPLIAPVLIIVSGDLILGW  
STYGLKNSPRVYVTEHMSKYKMKVYIGSPMDGTGMYKVFVAVGATVAIL  
TCALLVKKKCPRAHONGTFSKTEPLAPKNEFPAGLIDDEEVIYDEYEPLERGY  
CQEPEDVNTFGCAVVEGERALNFKSALIASADRIILANSGRNDYS"  
15398..16080  
/note="ORF9"  
/codon_start=1  
/protein_id="AAC5103.1"  
/db_xref="GI:1486495"  
translation="MPEKTSAGADAAAGNRPKSRNELIPLHRTGKTKTAGLSNY  
TOPIPNPKCSAGSDNHCADTDTYRRTCCASRYSQSDPSHPMPTETGVRPS  
AKRRKLSSDCEGAHQPLVSKLPDSQAPATYSQAPRYVDVSSPMPGVADAD  
LETRALPGATVEQTESKRLPNQOSHLKFTNHEVGECPCPEGIVEAPSLGLISR  
VGAIALNELARRRACLEPLAASAAAGIVAMAAALDQQR"  
16004..>18912  
/note="Large inverted repeat flanking the short unique  
region"  
/rpl_type=inverted  
16129..17013  
/note="coding region unique to avian herpesviruses"  
/codon_start=1  
/product="unknown"  
/protein_id="AAC5104.1"  
/db_xref="GI:1486496"  
translation="MSKCYCLARHLKSPCVRRVAFGLATMSRPTSLDLAFSA  
AFRGDLPGRFMRASQSDIFEPDLAIVQAAVAREGERISLOVAEDITNHD  
PRIAPAKRVAVAELGWLTALSELVSGPNEISKWKGOIPEPRAAMEIDVRYVI  
GPILSFESAVALPELVYIRAGGASRSRANMOSAPAGLAIRIGEMERSLV  
IALPLSFTLPEDLEGSQNSIRAFVHILNCAVATDKIMSDVRVPEESFYSCLRE  
IIMCERAPCPNPPKM"  
complement(17380..18216)  
/codon_start=1  
/product="US10"  
  
Query Match 99.9%; Score 1087.4; DB 14; Length 18912;  
Best Local Similarity 99.9%; Pct No. 4,5e-295;  
Matches 1088; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 ATGGCATGCTACTGGAACTGCTGCTGCTGCGCGGACGCTGCGACCTTGCGGCGG 60  
Db ATGSCATGCTACTGGAACTGCTGCTGCTGCGCGGACGCTGCGACCTTGCGGCGG 12569  
QY 61 ATGGGAATGCTGACTGGAATACGCTTCGCCGAGATTGACAGACATGACATGCTG 120  
Db ATGGGAATGCTGACTGGAATACGCTTCGCCGAGATTGACAGACATGACATGCTG 12629  
QY 121 ATCGTCGGCGCTGCGCCGGAAGCTACATGACGCTATTGTTCAATGCTGCGCAG 180  
Db ATCGTCGGCGCTGCGCCGGAAGCTACATGACGCTATTGTTCAATGCTGCGCAG 12689  
QY 12630 ATCGTCGGCGCTGCGCCGGAAGCTACATGACGCTATTGTTCAATGCTGCGCAG 240  
Db ATCGTCGGCGCTGCGCCGGAAGCTACATGACGCTATTGTTCAATGCTGCGCAG 12749  
QY 241 CAGTGTACACAGAACTAGCAGAGAGCGCTTGAAATTCATCATGATGATGCTCTCT 300  
Db CAGTGTACACAGAACTAGCAGAGAGCGCTTGAAATTCATCATGATGATGCTCTCT 12809  
QY 301 GTTTTGTGGGCTGAAAGTGACCGAGTACAGTTCCTCGGCTCGAAGACATACCGGA 360  
Db GTTTTGTGGGCTGAAAGTGACCGAGTACAGTTCCTCGGCTCGAAGACATACCGGA 12869  
QY 361 CTTCCACACCGCTTAAAGCTCAGATACGAAATTCCTCTCGAAGACACACGGGATGTT 420  
Db CTTCCACACCGCTTAAAGCTCAGATACGAAATTCCTCTCGAAGACACACGGGATGTT 12929
```

```
QY 421 TACGTAAATTTGTTGCGCTAGACAGACACCAAGAACCCATTGACGCTTTCGGATCCAACTA 480  
Db TACGTAAATTTGTTGCGCTAGACAGACACCAAGAACCCATTGACGCTTTCGGATCCAACTA 12989  
QY 481 TCGGTATATCAATTTGCGACACACCGCGGACGCTGCGGACCTATTTCAAGAGCTTGCTGT 540  
Db TCGGTATATCAATTTGCGACACACCGCGGACGCTGCGGACCTATTTCAAGAGCTTGCTGT 13049  
QY 541 CGCACCTTGCGATTACCTACCTGCTCAACCTTGAAGCCTATCTCAGAGCCGAGGAAATTTGG 600  
Db CGCACCTTGCGATTACCTACCTGCTCAACCTTGAAGCCTATCTCAGAGCCGAGGAAATTTGG 13109  
QY 601 CGCAACTGCGACCGCTAGCTTCCACAGGAGGCGCAGACGACCAAGCGCGGAGGACAAAC 660  
Db CGCAACTGCGACCGCTAGCTTCCACAGGAGGCGCAGACGACCAAGCGCGGAGGACAAAC 13169  
QY 661 CGGAGCGCGCGCTACCTGCAACAGCGCGCTCCGAACCTTGAAGCGGAACACTTACCTTCC 720  
Db CGGAGCGCGCGCTACCTGCAACAGCGCGCTCCGAACCTTGAAGCGGAACACTTACCTTCC 13229  
QY 721 TGGCTAGAAAATGCGGTGATCATAGCAACCGACACCGCCGAAAGAAATTCAAACGTT 780  
Db TGGCTAGAAAATGCGGTGATCATAGCAACCGACACCGCCGAAAGAAATTCAAACGTT 13289  
QY 781 ACTGTCGCTGCGGACCATGAGCCCTACGCTAATTTGGGTTAACCGTGGCTGCGCTGCTG 840  
Db ACTGTCGCTGCGGACCATGAGCCCTACGCTAATTTGGGTTAACCGTGGCTGCGCTGCTG 13349  
QY 841 AGCGACACATCGCGCTGCTCATTTGATTTTCATGCTACACGAAACATGTGACCCCG 900  
Db AGCGACACATCGCGCTGCTCATTTGATTTTCATGCTACACGAAACATGTGACCCCG 13409  
QY 901 CACGGAATTTAGACACGCTGCTGCAAGACGACGAAAGACGTTCCCAACATGAAAGGGA 960  
Db CACGGAATTTAGACACGCTGCTGCAAGACGACGAAAGACGTTCCCAACATGAAAGGGA 13469  
QY 961 TCGCGAAAATTTGGACCATGTTGCTGCTGCGCAATTAACAGGCGCGCTGACAGATAGT 1020  
Db TCGCGAAAATTTGGACCATGTTGCTGCTGCGCAATTAACAGGCGCGCTGACAGATAGT 13529  
QY 1021 GAACCTTGTAACATGTTGCGATTTTAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
Db GAACCTTGTAACATGTTGCGATTTTAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13589  
QY 1081 AAAATGTGA 1089  
Db AAAATGTGA 13598  
  
RESULT 2  
HSMDDGLYCO 1627 bp DNA linear VRL 06-MAY-1996  
LOCUS Gallid herpesvirus 1 glycoprotein D (gpD) gene, complete cds.  
DEFINITION L31965  
ACCESSION L31965  
VERSION L31965_1 GI:493595  
KEYWORDS glycoprotein D.  
SOURCE Gallid herpesvirus 1.  
ORGANISM Gallid herpesvirus 1.  
VIRUSES; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Infectious laryngotracheitis-like viruses.  
REFERENCE Johnson, M.A., Tyack, S.G., Pridaux, C.T., Kongsuan, K. and  
Shepard, M. Sequence characteristics of a gene in infectious laryngotracheitis  
virus homologous to glycoprotein D of herpes simplex virus  
JOURNAL DNA Seq. 5 (3), 191-194 (1995)  
MEDLINE 95337426  
PUBMED 7612933  
FEATURES  
location/Qualifiers  
1..1627  
/organism="Gallid herpesvirus 1"  
/strain="SA-2"
```

| FEATURES  | source   |
|---|--|
| gene  | /db xref="taxon:10386"<br>/note="also known as infectious barygottachetits virus (ILTV)"   |
| TATA_signal   | 144..1517<br>/gene="gd"  |
| CDS   | 144..151<br>/gene="gd"<br>326..1441<br>/gene="gd"  |
| polyA_signal  | /codon_start=1<br>/product="glycoprotein D"<br>/protein_id="AA98925.1"<br>/db_xref="GI:493596"<br>/translation="MDRHLFLRNAFMFTVLLSPASOSTAAVTVYITLGRALDALPT<br>PAPVPRNYLTVRSRGEIVLEINISNVDDMISAKEKRECPREASVSPYVTKGDD<br>GEDKCYPIYRKREYECGDVQLSCAQSOMALDVSSTLSVRNGAGLTIESPNAA<br>LSGGYLLFKIGRFAQTALVLENDKLTGTSNPLPSKCMWTEYQYTGEOEHLX<br>PIADTNRHADVDYRRYEDILQNNNNLLRKNPSAPRPDSVQELPAYTKAEGRT<br>PDASSERKAPPEDEDDMQAEASGEPMALPEDEDETHDDPNSDPDYNDMPRA<br>VPIVETTKSSNAVSMPIFAFVACAAVSMQVWGAS"<br>1507..1517<br>/gene="gd" |
| BASE COUNT  | 410 a 428 c 428 g 361 t  |
| ORIGIN  |  |
| Query Match   | 5.8%; Score 63; DB 14; Length 1627;  |
| Best Local Similarity                                       | 100.0%; Pred. No. 1.7e-06;   |
| Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |  |
| OY  | 1 ATGGCATCGCTACTGTGGAACCTCTGGCTCTCTTTCGCCGAGCGCTCGACACCTTGGCGCG 60<br>   |
| Db  | 1565 ATGGCATCGCTACTGTGGAACCTCTGGCTCTCTTTCGCCGAGCGCTCGACACCTTGGCGCG 1624  |
| OY  | 61 ATG 63<br>  |
| Db  | 1625 ATG 1627  |
| RESULT 3  |  |
| AF429315  | 125020 bp DNA linear PRI 18-JAN-2002   |
| LOCUS   | AF429315   |
| DEFINITION  | Homo sapiens junctophilin 3 (JPH3) gene, partial cds.  |
| DEFINITION  | AF429315   |
| VERSION   | AF429315.1 GI:17646244   |
| KEYWORDS  |  |
| SOURCE  | Homo sapiens.  |
| ORGANISM  | Homo sapiens<br>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.   |
| REFERENCE   | 1 (bases 1 to 125020)<br>Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,<br>Ingersoll,Ashworth,R.G., Flesher,A., Stevanin,G., Brice,A.,<br>Potter,N.T., Ross,C.A. and Margolis,R.L.<br>A repeat expansion in the gene encoding junctophilin-3 is<br>associated with Huntington disease-like 2<br>Nat. Genet. 29 (4), 377-378 (2001)   |
| TITLE   |  |
| JOURNAL   |  |
| MEDLINE   | 21583737   |
| PUBMED  | 11694876   |
| REFERENCE   | 2 (bases 1 to 125020)<br>Holmes,S.E., Ingersoll,Ashworth,R.G., Ross,C.A. and Margolis,R.L.<br>Direct Submission<br>Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical<br>Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA   |
| FEATURES  | location/Qualifiers<br>1..125020<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/chromosome="16"<br>/map="16q24.3: between D16S520 and W1-12410"<br>/note="Isolated from a patient with Huntington's<br>disease-Like 2 (HDL2)"<br>complement(35581..35746)<br>/rpt_type=tandem  |

[illegible]

|                       |  |   |     |        |                 |
|-----------------------|--|---|-----|--------|-----------------|
| RESULT 4              | AX164174   | 1246 bp   | DNA | Linear | PAT 22-JUN-2001 |
| LOCUS                 | AX164174   | Sequence 4 from Patent WO0138564.                                 |     |        |                 |
| DEFINITION            | AX164174   |   |     |        |                 |
| ACCESSION             | AX164174   |   |     |        |                 |
| VERSION               | AX164174.1   | GI:14545112   |     |        |                 |
| KEYWORDS              |  |   |     |        |                 |
| SOURCE                | human.   |   |     |        |                 |
| ORGANISM              | Homo sapiens   |   |     |        |                 |
| REFERENCE             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.        |   |     |        |                 |
| AUTHORS               | 1 (bases 1 to 1246)<br>Rouleau, G.A., Latreille, R.G., Rochefort, D., Cossette, P. and Ragsdale, D.                                |   |     |        |                 |
| TITLE                 | Locl for idiopathic generalized epilepsy, mutations thereof and method using same to assess, diagnose, prognosis or treat epilepsy |   |     |        |                 |
| JOURNAL               | Patent: WO 0138564-A 4 31-MAY-2001;  |   |     |        |                 |
| FEATURES              | Location/Qualifiers  |   |     |        |                 |
| source                | 1. 1246<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"   |   |     |        |                 |
| BASE COUNT            | 112 a 36 c 104 g 109 t 885 others  |   |     |        |                 |
| ORIGIN                |  |   |     |        |                 |
| Query Match           | 4.0%; Score 43.2; DB 6; Length 1246;   |   |     |        |                 |
| Best local similarity | 12.7%; Pred. No. 0.63;   |   |     |        |                 |
| Matches               | 39; Conservative 141; Mismatches 127; Indels 0; Gaps 0;  |   |     |        |                 |
| QY                    | 65 GAATCGTACCTGAGAAATGACGCTCCGCGGATGACGATGACATCATGATCG 124   |   |     |        |                 |
| DB                    | 445 SHKESBSSYVYKAAACNNHATMHCBSABSCBSBNCNBASDMCMCTPTXSSV 386  |   |     |        |                 |
| QY                    | 125 TCGGCGCTCCGCGGAGAGCTCAATCAATGACGATATTTTCACTGCGGAGAGAC 184  |   |     |        |                 |
| DB                    | 385 SASNSNTICDYBHHYSHNADHHTNHSCHMYTCSSAIVSYSSCSDSSSYVRAV 326   |   |     |        |                 |
| QY                    | 185 CCCCAAAACCTACTCAGACACCGCTCCGCGGCTGCTGATATACCAACGAGT 244  |   |     |        |                 |
| DB                    | 325 NCSYCMYTHSSMHMCCSMYMYNYMTSMSSSHSYCTSYSDSTATATTTT 266   |   |     |        |                 |
| QY                    | 245 GCTACCGAGAACTAGGAGAGAGCGCTTTGAATTCATCATCATGCTCTCTGTT 304   |   |     |        |                 |
| DB                    | 265 MMKTYATNRTKTBTRNSCBRRACCTTIANNNIAKSTMSHSAKCNKNTCTBGRK 206  |   |     |        |                 |
| QY                    | 305 TTGCGGCTGTAAGTACGAGCTACGCTTCCGCTCGAACAACACTAACCGACTC 364   |   |     |        |                 |
| DB                    | 205 GCTHSSNCGTCTCRDRYSHRSMHMBANACNRNANNSDSNTAMGNTNCKCTBSGAKHS 146  |   |     |        |                 |
| QY                    | 365 CACACCC 371  |   |     |        |                 |
| DB                    | 145 MMBSTCB 139  |   |     |        |                 |
| RESULT 5              | AP003261   | 138025 bp   | DNA | linear | PLN 10-AUG-2002 |
| LOCUS                 | AP003261   | Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, |     |        |                 |
| DEFINITION            | PAC clone: P0471B04.   |   |     |        |                 |
| ACCESSION             | AP003261   |   |     |        |                 |
| VERSION               | AP003261.3   | GI:22202658   |     |        |                 |
| SOURCE                | Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,   |   |     |        |                 |
| ORGANISM              | clone: P0471B04.   |   |     |        |                 |
| REFERENCE             | Oryza sativa (japonica cultivar-group)   |   |     |        |                 |
| AUTHORS               | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;   |   |     |        |                 |
| TITLE                 | Sphenolopsida; Magnoliophyta; Liliopsida; Poales; Poaceae;   |   |     |        |                 |
|                       | Eriophytaceae; Oryzaceae; Oryza.   |   |     |        |                 |
|                       | 1 Sasaki, T., Matsumoto, T. and Yamamoto, K.   |   |     |        |                 |
|                       | Oryza sativa (japonica) (G3) genomic DNA, chromosome 1, PAC  |   |     |        |                 |
|                       | clone: P0471B04  |   |     |        |                 |

## JOURNAL

Published Only in Database (2001)  
2 (bases 1 to 138025)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (19-FEB-2001) Takui Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasakienias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
On Aug 9, 2002 this sequence version replaced gi:21202837.  
Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLAST2.0, BLASTX2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against  
NCBI Nonredundant Protein database, nr  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RGP. Protein homologies of the coding regions were searched against  
NCBI Nonredundant Protein database with BLAST2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DBI accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
'unknown protein'. A gene predicted with a gene prediction program  
is classified as a 'hypothetical' protein.  
The orientation of the sequence is from SP6 to T7 of the PAC clone.  
This sequence of P0471B04 clone has an overlap with P0698A10 clone  
(DBJ: AP003287) at the position 1 to 24,438 of 5' end and an  
overlap with P0018C10 clone (DBJ: AP003227) at the position 48,801  
to 138,025 of 3' end. Detailed information on overlap and assembly  
quality together with annotation of this entry is available at  
http://rgp.dna.affrc.go.jp/Genomeseg.html.

## FEATURES

## source

1. 138025  
/organism="Oryza sativa (japonica cultivar-group)"

/cultivar="Nipponbare"  
/db\_xref="taxon:39947"

/chromosome="1"

/clone="P0471B04"  
/join(9533..9715,9887..10019,10850..10988,11365..11401,  
13640..13931,14823..14955,15564..15576)

/gene="P0471B04.1"  
/join(9533..9715,9887..10019,10850..10988,11365..11401,  
13640..13931,14823..14955,15564..15576)

/gene="P0471B04.1"  
/note="hypothetical protein  
similar to Arabidopsis thaliana chromosome 1, F10C21\_14"

/codon\_start=1  
/protein\_id="BAC07317.1"

/db\_xref="GI:22202659"

/translation="MSPPEVIGQFGDTTYRVEVGIAWETQETMRKREYRGEGLIE  
AVITDKNTRGSKGYGFTRPDPAAMACVADPAVIDGRANCMASLGVORRPT  
POHGARSFRVMSFSQOAGIGGAGAPSHATPFHVAIPQGLYHYGVSPYSDY  
SYPTNYNYITGAGVOYFPGAAAMAAAGVWSSPFYFPOFGSGSTTNYASGO  
XNIOYPOWHPFSTVASTAAVATGPAOQYGGPLSLAASPQAAAKRAGKSNPNMAL  
TGGCFMLIONMRTKYSKQIIEQCHFEISRP"

/complement(join(117199..17342,17914..18008,18141..18178,  
18784..19073))

/gene="P0471B04.2"

/complement(join(117199..17342,17914..18008,18141..18178,  
18784..19073))

/gene="P0471B04.2"

/product="putative SnRK1-interacting protein 1"

/protein\_id="BAC07318.1"

/db\_xref="GI:22202660"

/translation="MASGALAKLILGRRAATPTLLAPPAARARSRPPOAPAPSSD  
EDDFAGGVAAPATGISTAPLAEVRLGRGKRVENLYTRDPEPSIKYIPMHIVK  
IINIAPEWSEVRSIVYSSDKSVSYIIRTLTGTAETIYRACGTSSADDTGCGP  
YOKAAMAFERRACRGLGLHLEHEDMS

/join(24593..24925,25029..25403)

/gene="P0471B04.3"

gene



JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

clone:P0018C10  
Published Only in Database (2001)  
2 (bases 1 to 200183)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (19-FEB-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsusakienias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
On Jul 24, 2002 this sequence version replaced g1:21202836.  
GENSCAN 0.0, BLASTX 2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against  
NCBI Nonredundant Protein database, nr  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RGP. Protein homologies of the coding regions were searched against  
NCBI Nonredundant Protein database with BLASTP 2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DBJ accession no. and RGP clone ID.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, putative, and "like protein". A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
"unknown" protein. A gene predicted with a gene prediction program  
is classified as a "hypothetical" protein.  
The orientation of the sequence is from Sp6 to 77 of the PAC clone.  
This sequence of P0018C10 clone has an overlap with P0471804  
clone (DBJ: AP003261) at the position 1 to 89,225 of 5' end and an  
overlap with B1055E10 clone (DBJ: AP003561) at the position  
158,410 to 200,183 of 3' end. Detailed information on overlap and  
assembly quality together with annotation of this entry is  
available at <http://rgp.dna.affrc.go.jp/genomeseg.html>.

## FEATURES

## source

1. 200183  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="1"  
/clone="P0018C10"  
join(6052..6159,6257..6388,6483..6612,6719..6805,  
7232..7302)  
/gene="P0018C10.1"  
join(6052..6159,6257..6388,6483..6612,6719..6805,  
7232..7302)  
/gene="P0018C10.1"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC06200.1"  
/db\_xref="GI:21952786"  
/translation="MTISLARSSTCTGTADTAPAKPAGAPOLKILADISLE  
EFDSICPPICTGELTLPDSNFAVYTDPLLYQVTRKCGFVSSTQICHNLVDA  
AGTQFLQAVGELVGAERFAGDELVHRKFLSGPDVVALRDQLPTLSFGCHVSKTK  
QKCYQETGEWERY"  
complement(join(14216..14383,14464..14772))  
/gene="P0018C10.2"  
complement(join(14216..14383,14464..14772))  
/gene="P0018C10.2"  
/note="contains EST A0058082(E30867)  
unknown protein"  
/codon\_start=1  
/protein\_id="BAC06201.1"  
/db\_xref="GI:21952785"  
/translation="MSGSRPMPCDDPSPAPQPPVVAANAAGCASTLNDFTGSD  
AISFGFAVAILSMELMAIFELIRKPHYFPLASALRLPARRHGVSPGKLRPP  
MVEVLOADLVLMGQRPYLAQAPLPAPPCREGVHMPRDHDVHNSVMP"  
complement(join(21230..21448,21542..21620,22147..22244))  
/gene="P0018C10.3"  
complement(join(21230..21448,21542..21620,22147..22244))  
/gene="P0018C10.3"  
/note="hypothetical protein"  
/codon\_start=1

gene  
CDS

/protein\_id="BAC06202.1"  
/db\_xref="GI:21952786"  
/translation="MGYGRILSLRSLVTSLRNLQNPDTNLHQNNKTRHADVETDLP  
KKNKIASNVASAEKREAKKNGKRRRPPGVSIRPPVHHGVAAASRRSSNPLGDEL  
DPLTMSERAVHLPAREKPVKASIRPSP"  
join(26342..28883,29797..30149)  
/gene="P0018C10.4"  
join(26342..28883,29797..30149)  
/note="contains ESTs  
C96730(C10643),C2379(C12740),C22380(C12740)"  
/codon\_start=1  
/product="putative receptor protein kinase"  
/protein\_id="BAC06203.1"  
/db\_xref="GI:21952787"

## gene

## CDS

/translation="MLPQDQIYICFILLSLKRTISLPLEFDALDIKSHEDPON  
YIGNMDESHPQCYGYTCDDYSGVIGTISLNSASISGTSISFSLSQRLTEGNA  
SISGIIIPALANCLNLOVNLSTNSLTQDLSTFTNLQDLDTNNNSGFFPANNV  
KLSGILTEIGLENNFNENGDPSIKELNLTWLPFLNGLRGLPVALRDLVSLGTLID  
FERNQITGVPEPIAISNLRLMKTELYQNNLTGETLPPLALITLLSEFDVSNQGRSLP  
PKEIANLKKIKFHHYNNNSGVLEGGDLDEFSEFTEENQSPKRPANLGRSPDL  
NAIDSENVFSGEPPEPLCNMKNLQFLATLNNFSGEPSSYSCKTQDRETIQNDP  
TORHSGIMGJPNNAVILDAVNAVMSFATISDQISASLNLQVYNNVNSGELPDELK  
LSLQKVAFNRRSGQIPAQIGSLKQSLFHLBQNALIEGSIIPDQIMCNSLVDLNLA  
DNLGTGIPDLASTFLNLSLNSHMTISGETIPGQLYKLISYDFSHNLSGVPPA  
LMIAGDPAFSENDGLCIAGVSEGMQATNLRYCPWNNDHQNFSORRLVVLIVYS  
LVLLSGLACIARENYKLEQFHSKQDIEGSDDSKMYLTFPRLDPEELCNLDLAF  
NLGGCGTGVYRLELSKGRVYAVQMLKRDADKVRTEINTGIRRRNTLKHAF  
LIGSENLVEYVNGNLDAIRREFKQPELDMKRYRIVGTAKQIMVLIHDS  
PATIRDRKSTNILLDEYKALADGKILVSGPLSCFAGTGVABELAVSLKVT  
EKSDYSGIYLLLTGRSPQCPGDLIDVSVSHLANQNPAAVLDPVSSHAS  
EMTWTVALICIVQLPSEKPMREVMMLDIDISANGKKNNDK"  
complement(join(30791..31219,31321..31545,31728..32088,  
32199..32292,32370..32603,33125..33179))  
/gene="P0018C10.5"  
complement(join(30791..31219,31321..31545,31728..32088,  
32199..32292,32370..32603,33125..33179))  
/gene="P0018C10.5"  
/codon\_start=1  
/product="putative amino acid permease 6"  
/protein\_id="BAC06204.1"  
/db\_xref="GI:21952788"

gene  
CDS

/translation="NNKKAPEDEYEGSEHETGYVTAATVAVTGGVLAIAVSV  
AOLGVAVGLAAGFACVYITSTLANNKRPAPHYTTRRRYMDNVSYSISPEVF  
MCGIAOVNLMGTWYGTITATVNAIRNSOCHRRNGAANACNTSAYULMAFS  
IVQVILSOPGLEHTWLSYVAVMSFATISDQISASLNLQVYNNVNSGELPDELK  
SSSKLLMNVLLAGNIAFAVLEIOTDLKSPENKTRKANKAGTCATATTFY  
ISVGCAGVAAFGSDAPGNITLASCQPEMDVAPYLVLTIVTAVTAVVAMIPFNK  
RMISNRMPAKFINSEYVNPVLIQKSTVAPYLVLTIVTAVTAVVAMIPFNK  
VLGLGAFSEFMDIYVFPISMHIAQKIRGRNWLILGLSMVCLMISVAVGIGVTD  
IVSLKAVAPFKTVS"  
join(43546..43657,43872..43972)  
/gene="P0018C10.6"  
join(43546..43657,43872..43972)  
/gene="P0018C10.6"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC06205.1"  
/db\_xref="GI:21952789"  
/translation="MRSIITLSCFDKYSKGIHYTSSQGRPLPYAAKATGASAEKFF  
KIVELELEINVTYRKWKMLGYLN"  
join(44656..44752,44835..45020,45082..45407,45609..45665)  
/gene="P0018C10.7"  
join(44656..44752,44835..45020,45082..45407,45609..45665)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC06206.1"  
/db\_xref="GI:21952790"

gene  
CDS

/translation="MKGVEGRGRSGSGASLTSGAGWGTGSGADAVTGHRRRAG  
TQGNDAVEEGVAGARELVTKGAVEEBSVYNAKQDDEPRGCGVRRAGALGGY  
RRRGRGRNGKQDGLPLAVRKYVPASTHRSCTHRSACPLHLPRAAVALSS  
RPLPAPSGCRSACAPLRPCTCSAPRAFSRRKRPALNDPLVILVNLKQEN



[illegible]

| FEATURES              | source   | repeat_region | gene | mrna | cds | BASE COUNT | ORIGIN |
|-----------------------|--|---------------|------|------|-----|------------|--------|
| JOURNAL               | associated with Huntington disease-like 2  |               |      |      |     |            |        |
| MEDLINE               | Nat. Genet. 29 (4), 377-378 (2001)   |               |      |      |     |            |        |
| PUBLISHED             | 21583737   |               |      |      |     |            |        |
| REFERENCE             | 2 (bases 1 to 125020)  |               |      |      |     |            |        |
| AUTHORS               | Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.  |               |      |      |     |            |        |
| TITLE                 | Direct Submission  |               |      |      |     |            |        |
| JOURNAL               | Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA |               |      |      |     |            |        |
| source                | Location/Qualifiers  |               |      |      |     |            |        |
|                       | 1. 125020  |               |      |      |     |            |        |
|                       | /organism="Homo sapiens"   |               |      |      |     |            |        |
|                       | /db_xref="taxon:9606"  |               |      |      |     |            |        |
|                       | /chromosome="16"   |               |      |      |     |            |        |
|                       | /map="16q24.3: Between D16S520 and WI-12410"   |               |      |      |     |            |        |
|                       | /note="Isolated from a patient with Huntington's disease-Like 2 (HDL2)"  |               |      |      |     |            |        |
|                       | complement(35581..35746)   |               |      |      |     |            |        |
|                       | /rpt_type=tandem   |               |      |      |     |            |        |
|                       | /rpt_unit=ctg  |               |      |      |     |            |        |
|                       | complement(<36507..>36887)   |               |      |      |     |            |        |
|                       | /gene="JPH3"   |               |      |      |     |            |        |
|                       | /note="Jp3"  |               |      |      |     |            |        |
|                       | complement(<36507..>36887)   |               |      |      |     |            |        |
|                       | /gene="JPH3"   |               |      |      |     |            |        |
|                       | /product="junctionophlin 3"  |               |      |      |     |            |        |
|                       | complement(<36507..>36887)   |               |      |      |     |            |        |
|                       | /gene="JPH3"   |               |      |      |     |            |        |
|                       | /note="component of the junctional complex between plasma membrane and endoplasmic reticulum"                      |               |      |      |     |            |        |
|                       | /codon_start=1   |               |      |      |     |            |        |
|                       | /product="junctionophlin 3"  |               |      |      |     |            |        |
|                       | /protein_id="AAL0941.1"  |               |      |      |     |            |        |
|                       | /db_xref="GI:17646245"   |               |      |      |     |            |        |
|                       | /translation="MSGSGRFRNFDGSGYCGWEDGKAHGHVCTGPKGCGEYTGSM  |               |      |      |     |            |        |
|                       | HGFELVLTWPSGNTQGTGTAAGKRGIGLIESKGVVYKGEWTHGPKRGYGVRECA   |               |      |      |     |            |        |
|                       | NGKYEKSTMSGNGLDGCTEYTSIDG"   |               |      |      |     |            |        |
| Query Match           | 3.7%; Score 40.2; DB 9; Length 125020;   |               |      |      |     |            |        |
| Best Local Similarity | 10.7%; Pred. No. 9.9;  |               |      |      |     |            |        |
| Matches               | 33; Conservative 144; Mismatches 132; Indels 0; Gaps 0;  |               |      |      |     |            |        |
| Y                     | 171 GCCTGCCAGAGACCCACAACCCCTACGAGAACCGTCGCGTTCGGTGTGA 230  |               |      |      |     |            |        |
| Db                    | 71647 SCWMSMMAAAYMKCGMSAMCYYKKYWCYACMAAAYMMMAATYARSMGASMGWK 71581  |               |      |      |     |            |        |
| Y                     | 231 TATAACAACAGCGTCAACAGAGAACTTACGAGAGCGCTTGAATAATGCACTATCG 290  |               |      |      |     |            |        |
| Db                    | 71587 YYSAMSWCCWSMMMYSMNGCRMRMRMSMSMRRSCMRASIMYRSMWKAACCT 71522  |               |      |      |     |            |        |
| Y                     | 291 ATCGTCTTCTGTTTTTCTCGCTGTAAAGACCGAGTACGCTTCTCCGCTCGAAG 350  |               |      |      |     |            |        |
| Db                    | 71527 RRRASSTRRRRRKYSMAYSARMMRMAMWTAMRHCYIYSAACMCNCKMYISGCRMCNR 71466  |               |      |      |     |            |        |
| Y                     | 351 ACTAACCGGACCTCCACACCCGTTTAAAGTCACTATACGAATCTCTGTCGGAAGAC 410   |               |      |      |     |            |        |
| Db                    | 71467 RMAAAMMYUAYMYKAYARMAMRABARMAMARBRMRMMMKKTKTKKYMMWSRSSWCAC 71408  |               |      |      |     |            |        |
| Y                     | 411 CGGATCTTCTAGCTAATTTGTTGGGCTAGACGACACCAAGAACCACTTACGCTTTCCG 470   |               |      |      |     |            |        |
| Db                    | 71407 RCYIRMMYMMMSMSAMTKWRRRARARSCCRASSGMSMSAYMMMSWSSWCMGRMSWYRA 71348   |               |      |      |     |            |        |
| Y                     | 471 GATCCAACT 479  |               |      |      |     |            |        |
| Db                    | 71347 GACCATCT 71339   |               |      |      |     |            |        |
| RESULT 8              | DMU19909   |               |      |      |     |            |        |
| LOCUS                 | DMU19909   |               |      |      |     |            |        |
| DEFINITION            | Drosophila melanogaster coxscrew protein (csW) gene, complete cds.   |               |      |      |     |            |        |









gene /db\_xref="GI:13880052"  
/translation="MGCGITICLNTSEVPNNENRKNKNOAGLDSIRVIRHSPFDIDPE  
PDSGIDVYWSQDAIILHAPDRKRVLEEARVLRPGCELLFTDPMADVDVDPGVLDPVTLT  
RLNLRDLSGMRPYA"  
3948..4253  
/gene="MT05433"  
3948..4253  
/note="Identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44766.1"  
/db\_xref="GI:13880053"  
/translation="MREAQALGFEEVLDRDLVRLRPHYSRVFEELEFARLEIEGKSS  
SESEYLDKMKVYSGKNVNEADNGSHVGHPTPEPRRLPICPLPAIDSTAGRRRYR  
4388..5692  
/gene="MT05444"  
4388..5692  
/note="similar to SP:P46349 PID:969027 PID:1708639  
GG:AL009126; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="amino acid permease"  
/protein\_id="AAK44767.1"  
/db\_xref="GI:13880054"  
/translation="MAIAGVIGAGLFEVSGVIRATGPAFLTYALGALIVLMRR  
LGEAANSTGAFADYDAKALAGWAGFSGVLMYFHWIYVGEVAVAGVGLTYWIDIE  
ALPMILSLGKMMATNLSVSSGFEFMDAGYKVTITGPIVLGAPFGLDPPGH  
GMDPNSLCHMGFPDGGVGAFAAIVVAIFEMGTGVEYTTIAARPDQRAVQASST  
VYARIVPFGSVFLITVILPNRSLDELASPYVALRMRIGCGADQIMNAVLRATLS  
CLNSELINSRMLFVLAKRQEPALVLRNKGPTFAIMGSSVGEGLIMAVNSPRA  
TVFVLELNSGALFVLEVLIALLSQIVLRQTSQNLGVRMLFPGSLIYVTVIVAVN  
LARMFEDLAARQQLSLISMAVVGCVLTVTLVRRPLRPW"  
complement(5676..6080)  
/gene="MT0545"  
complement(5676..6080)  
/note="similar to GB:U00018 SP:P46716 PID:467041;  
/gene="MT0545"  
/note="Identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44768.1"  
/db\_xref="GI:13880055"  
/translation="VWGMQLPQMLARFNRYVNTPIQRLMAGMLPAILHEVGRSGRK  
PRTPLNFSADVDRAGVAAILITGPRMDLKNKTAGCGMRMRGCTFCVANDRRLT  
TKEAPAYVSSKWRVRFARLPDEAVLITKAD"  
6185..7573  
/gene="MT0546"  
6185..7573  
/note="similar to GB:U00018 SP:P46716 PID:467041;  
/gene="MT0546"  
/note="Identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate-1-semialdehyde 2,1-aminomutase"  
/protein\_id="AAK44769.1"  
/db\_xref="GI:13880056"  
/translation="MGSEFQATSRVIRGAARTSAQLEACSVIRGCVNPRAPTAG  
GTPRRTTAHGMLIDADGKNRVDLVCSGMGMITLGHAPHAYEVAKAAARGLSFGAP  
TPEALQLAGEITIGRVAPERIRLVNSGTETAVMSVRLARGTSGAKITGSGCGHGA  
DAALDSSGVAITGLCDPDRPASPRSSGSRGLSSGCVGAAGAADTIVLPYNDHGA  
VQSTFARGEGDIAAIVTEASPGNMGVPPGFPNALRAITAEIGALLIDDEVNTGRR  
VRSRWGVIDPVPADLFAFGKVMGSGMPAAAFGGRAEVMORLIGVYVQAGTISGNDP  
VAVAGIATLIRAADDAVYATLADANDRACLGLSALUDAVYPHOISRAGNNLSFSEE  
TPYVDFASARSSQTRIPYAFPHAMLDGVIYPPCSAFEMFVSALDDAAFGRIANALP  
AAARAAGERP"  
7573..8181  
/gene="MT0547"  
7573..8181  
/gene="MT0547"  
/note="Identified by match to protein family HMM"

|                          |  |   |
|--------------------------|--|---|
|                          | gene   | /codon_start=1<br>/transl_table=1<br>/product="phosphoglycerate mutase family protein"<br>/protein_id="AAK4770.1"<br>/db_xref="GI:13880057"   |
| CDS                      |  | /translation="MPETGYVHVNRHGEVHNPTGILYGRLPFHLSTAGQAQAAAAYADALARDIYAIVASPIOROGTAAPLRAITDAVTDDPLIESAFPGRRYPGGDGAARDPFRWMLRNRPFTPSNGEPIVDIAARRTAVDKARRRGAGEHYVCVSQQLPWTLRLLTGRLRHMDPRRDCALASTSLYTGDRLDVVYSOPPAL"                  |
|                          |  | 8196..8846<br>/gene="MT0548"<br>8196..8846<br>/note="Identified by GLIMMER2; putative"<br>/codon_start=1<br>/transl_table=1<br>/product="conserved hypothetical protein"<br>/protein_id="AAK4771.1"<br>/db_xref="GI:13880058" |
|                          |  | /translation="MQSRATRSRALTRRLVIIVAAYSALLITGCSDRDVAOQGTEEFVSPSGKTDFIDPPPASNGRGPISGLPADPARSVSLDEPDGYYVVNWMGWCMCGPC  |
| Query Match              | Best Local Similarity  | 3.5%; Score 38.2; DB 1; Length 16984;   |
| Matches                  | Conservative   | Mismatches 48; Indels 0; Gaps 0;  |
| Df                       | 9386   | TCCCTGCCGACACGTTCGACCCTTTGGCGCGCATGGAGAATCATTGATCACTGGAAATCACG 88   |
| Oy                       | 89   | TCTCGCCGACGATTGACGACGATCATCATCTGATCTGTGGCCCCCTCGCCCACAAC 143  |
| Df                       | 9326   | TCGGCCCCGACGACTGTCCACGCCGACGCGAACCACCGCCCAAACACGCGCGC 9272  |
| RESULT 13 MSGY224/C      |  |   |
| LOCUS                    | MSGY224  | 40051 bp DNA linear BCT 03-DEC-1996   |
| DEFINITION               | Mycobacterium tuberculosis sequence from clone y224.   |   |
| ACCESSION                | AB000004   |   |
| VERSION                  | AD000004.1   | GI:1702969  |
| KEYWORDS                 |  |   |
| SOURCE                   |  |   |
| ORGANISM                 | Mycobacterium tuberculosis (clone: Y224) ds-DNA.<br>Mycobacterium tuberculosis<br>Bacteria; Firmicutes; Actinobacteria; Actinomycetales;<br>Actinomycetiales; Corynebacterineae; Mycobacteriaceae;<br>Mycobacterium; Mycobacterium tuberculosis complex.<br>1 (bases 1 to 40051) |   |
| REFERENCE                | Du,L.<br>Direct Submission<br>Submitted (11-OCT-1996)<br>Beaver Street,<br>JOURNAL Waltham, MA, USA, 02154 du@rlc.com<br>GBS:S:1004713   |   |
| AUTHORS TITLE            |  |   |
| JOURNAL COMMENT FEATURES |  |   |
| FEATURES                 | Location/Qualifiers<br>. . . . .<br>1..40051 /organism="Mycobacterium tuberculosis"<br>/db_xref="taxon:1773"<br>/clone="Y224"  |   |
| BASE COUNT               | 6264 A 13435 C 13774 G 6578 T  |   |
| ORIGIN                   |  |   |
| Query Match              | Best Local Similarity  | 3.5%; Score 38.2; DB 1; Length 40051;   |
| Matches                  | Conservative   | Mismatches 48; Indels 0; Gaps 0;  |
| Df                       | 4021   | TCCATGCCGACGACTGCACCCCTTGGCGCGCATGGAGAATCATTGATCACTGGAAATCACG 88  |
| Oy                       | 89   | TCTCGCCGACGATTGACGACGATCATCATCTGATCTGTGGCCCCCTCGCCCACAAC 143  |
| Df                       | 3961   | TCGGCCCCGACGACTGTCCACGCCGACGCGAACCACCGCCCAAACACGCGCGC 3907  |

RESULT 14  
LOCUS MY25D10/c 40838 bp DNA linear BCT 03-AUG-2001  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 28/162.  
ACCESSION Z59558 AL123456  
VERSION 295558.1 GI:3261781  
KEYWORDS  
SOURCE  
ORGANISM Mycobacterium tuberculosis H37Rv.  
Mycobacterium tuberculosis H37Rv.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE  
AUTHORS 1 (bases 1 to 40838)  
Cole, S.T., Brosch, R., Parkhill, J., Garner, T., Churcher, C.,  
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,  
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
Comor, R., Davies, R., Devlin, K., Fellwell, T., Gentles, S.,  
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
Squires, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
Barrell, B.G.  
TITLE  
Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence  
JOURNAL Nature 393 (6685), 537-544 (1998)  
MEDLINE 98295987  
PUBMED 9634230  
REFERENCE 2 (bases 1 to 40838)  
Parkhill, J.  
AUTHORS  
TITLE  
Direct Submission  
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
COMMENT  
Notes:  
Details of M. tuberculosis sequencing at the Sanger Centre are  
available on the World Wide Web  
(URL: [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have  
been renumbered from the original cosmid submissions but the old  
gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a hidden Markov Model of 18 genes  
implemented in TParse (Krogh) supplemented with visual inspection  
of positional base preference in codons, especially where there is  
an increase in the observed/expected third position G + C.  
CAUTION: In some cases we may not have predicted the correct  
initiation codon. Where possible we choose an upstream ribosome  
(acc, gta, or tta) which is preceded by an upstream ribosome  
binding site sequence (optimally 5-13bp before the initiation  
codon). If this cannot be identified we choose the most upstream  
initiation codon.  
FEATURES  
SOURCE  
1. 40838  
/organism="Mycobacterium tuberculosis H37Rv"  
/strain="H37Rv"  
/db\_xref="taxon:83332"  
<1. 480  
/organism="Mycobacterium tuberculosis H37Rv"  
/strain="H37Rv"  
/db\_xref="taxon:83332"  
/clone="Y20G10"  
1. 40095  
/organism="Mycobacterium tuberculosis H37Rv"  
/strain="H37Rv"  
/db\_xref="taxon:83332"  
/clone="Y25D10"  
/gene="Rv0523c"  
/note="Rv0523c, (MTCY25D10.02, unknown), len: 131 aa,  
source  
1. 480  
/organism="Mycobacterium tuberculosis H37Rv"  
/strain="H37Rv"  
/db\_xref="taxon:83332"  
/clone="Y25D10"  
/gene="Rv0523c"  
/note="Rv0523c, (MTCY25D10.02, unknown), len: 131 aa,  
gene  
/gene="Rv0523c"  
/note="Rv0523c, (MTCY25D10.02, unknown), len: 131 aa,  
CDS  
/gene="Rv0523c"  
/note="Rv0523c, (MTCY25D10.02, unknown), len: 131 aa,  
/note="Rv0523c, (MTCY25D10.02, unknown), len: 131 aa,

misc-feature  
1616 1726  
/gene="hemL"  
/note="ps00600 Amino transferases class-III  
pyridoxal phosphate attachment site"  
2209 .2817  
/gene="Rv0525"  
2209 .2817  
/gene="Rv0525"  
/note="Rv0525, (MTCY25D10.04), len: 202 aa, unknown,  
highly similar to M. leprae Q49821 B2168\_C3-276 (202 aa)  
opt: 1151 z-score: 1349.4 E(): 0, (82.5% identity in 200  
aa overlap), also similar to SLD0395 SYNECHOYSTIS SP.  
Q55734 hypothetical 23.8 kDa protein (212 aa) opt: 207  
z-score: 251.8 E(): 5.1e-07 (28.2% identity in 195 aa  
overlap), also some similarity to Y019\_MyCTV Q10512  
hypothetical 39.2 kD protein cy427.09 (364 aa) opt: 236  
z-score: 282.0 E(): 1.1e-08, (34.3% identity in 198 aa  
overlap)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv0525"  
/protein\_id="CA08992.1"  
/db\_xref="GI:2113987"  
/translation="MPEOVYVYRHSVNPVPTGILYGRPLPGFHLSTATGANOAAVAD  
ALADRIDVAVIASPIQAOEETAPLAAHPLAVETPDILISANFEERRRPGDGM  
RDPVWMOALNPFTSPMCEPVYDIAAMTAYAVKARVARGHEVVCVSHQLPVTLT  
YLCKRLMDPRRDCALASTSLYGDRLVDVYVYSGNAL"  
2832 3482  
/gene="Rv0526"  
2832 3482  
/gene="Rv0526"  
/note="Rv0526, (MTCY25D10.05), len: 216, unknown, highly  
similar to M. leprae Q49816 U218C (216 aa) opt: 1144  
z-score: 1259.4 E(): 0, (78.5% identity in 214 aa  
overlap), some similarity to thiodoxins eg RESA\_BAC50  
some similarity to M. tuberculosis MTCY336\_06 006592 (136  
aa); fasta scores, opt: 197 z-score: 236.8 E(): 5e-08,  
(38.4% identity in 99 aa overlap)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv0523c"  
/protein\_id="CA08990.1"  
/db\_xref="GI:2113985"  
/translation="MOLPOMTARPRYVPTPIORLWAGLPAPALIEHVGSRKPYR  
TPINVSADYDGRACVAILLYGRNPMINITIAAGGMRKRGTEFVAPNRLTITA  
EAPYVSSRRRPFYATLPDEAVLITRAD"  
821 .2209  
/gene="hemL"  
821 .2209  
/gene="hemL"  
/note="Rv0524, (MTCY25D10.03), len: 462, hemL, highly  
similar to GSA\_MYCLE P46716 glutamate-1-semialdehyde  
2,1-aminomutase (EC 5.4.3.8) (GSA)(446 aa), fasta scores,  
opt: 1532 z-score: 1581.4 E(): 0, (82.6% identity in 460  
aa overlap); contains PS00600 Amino transferases class-III  
pyridoxal-phosphate attachment site"  
/codon\_start=1  
/transl\_table=11  
/product="hemL"  
/protein\_id="CA08991.1"  
/db\_xref="GI:2113986"  
/translation="MGTFOATSRVRCGARTSAQLFEACSVTPGVNSPVRAFTAVG  
GTPREITEAHGCMILIDAGNRVYDVCSPMILIGHAHPVAVAVAKAAGLSEFAP  
TPAETOLAGIIGRAVAPERIRLVNLSGTEMTAVAVLAFGFRKAYKFGCGHGV  
DALLADGSGVATIGLCDPDRPSPSSSRPLPSPPTVTAADITVTPYDIDA  
VQOFAPRGSOVLAAYTETASRGNGVYPPGPNALNRTAEHALLIDETVGTGR  
VSRGWDGIDPPADLPFGVNGSGGAPAFGRVWQRLAPGYPYOGCTISGAP  
VAVAGATITRAADAVYATLADANADLALLSEALTVDVHVISRAGNNLSVFPE  
TPYDTPASARISQTRVPAFPAHMLDNGVYPCSAEAMFVSAALDDAAPFRIANALP  
AAAPAAAOEPPA"  
1616 1726  
/gene="hemL"  
/note="ps00600 Amino transferases class-III  
pyridoxal phosphate attachment site"  
2209 .2817  
/gene="Rv0525"  
2209 .2817  
/gene="Rv0525"  
/note="Rv0525, (MTCY25D10.04), len: 202 aa, unknown,  
highly similar to M. leprae Q49821 B2168\_C3-276 (202 aa)  
opt: 1151 z-score: 1349.4 E(): 0, (82.5% identity in 200  
aa overlap), also similar to SLD0395 SYNECHOYSTIS SP.  
Q55734 hypothetical 23.8 kDa protein (212 aa) opt: 207  
z-score: 251.8 E(): 5.1e-07 (28.2% identity in 195 aa  
overlap), also some similarity to Y019\_MyCTV Q10512  
hypothetical 39.2 kD protein cy427.09 (364 aa) opt: 236  
z-score: 282.0 E(): 1.1e-08, (34.3% identity in 198 aa  
overlap)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein Rv0525"  
/protein\_id="CA08992.1"  
/db\_xref="GI:2113987"  
/translation="MPEOVYVYRHSVNPVPTGILYGRPLPGFHLSTATGANOAAVAD  
ALADRIDVAVIASPIQAOEETAPLAAHPLAVETPDILISANFEERRRPGDGM  
RDPVWMOALNPFTSPMCEPVYDIAAMTAYAVKARVARGHEVVCVSHQLPVTLT  
YLCKRLMDPRRDCALASTSLYGDRLVDVYVYSGNAL"  
2832 3482  
/gene="Rv0526"  
2832 3482  
/gene="Rv0526"  
/note="Rv0526, (MTCY25D10.05), len: 216, unknown, highly  
similar to M. leprae Q49816 U218C (216 aa) opt: 1144  
z-score: 1259.4 E(): 0, (78.5% identity in 214 aa  
overlap), some similarity to thiodoxins eg RESA\_BAC50



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 13:03:03 ; Search time 324 Seconds  
(without alignments)  
7569.217 Million cell updates/sec

Title: US-08-468-190c-1\_COPY\_9874\_10962

Perfect score: 1089  
Sequence: 1 ATGCGATCGCTACTTGAGAC.....CCGACTCAATTAATAATGTA 1089

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

N.Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID | Description |
|------------|--------|-------------|---------|-------|-------------|
| 1          | 1089   | 100.0       | 13473   | 16    | AAT33504    |
| 2          | 1089   | 100.0       | 13473   | 17    | AAT44384    |
| 3          | 1087.4 | 99.9        | 18912   | 17    | AAT44385    |
| 4          | 346    | 31.8        | 534     | 16    | AAT33513    |
| 5          | 38.2   | 3.5         | 4403765 | 22    | AA199683    |
| 6          | 38.2   | 3.5         | 4411529 | 22    | AA199682    |
| 7          | 38     | 3.5         | 2526    | 22    | AA26455     |
| 8          | 38     | 3.5         | 2677    | 23    | AB106927    |
| 9          | 38     | 3.5         | 2838    | 23    | AB18521     |

|    |      |     |       |    |          |
|----|------|-----|-------|----|----------|
| 10 | 38   | 3.5 | 4528  | 23 | AB121135 |
| 11 | 38   | 3.5 | 10587 | 23 | AB18520  |
| 12 | 38   | 3.5 | 19468 | 23 | AB106926 |
| 13 | 38   | 3.5 | 21399 | 23 | AB121134 |
| 14 | 37.6 | 3.5 | 813   | 24 | ABO13968 |
| 15 | 37.6 | 3.5 | 813   | 24 | ABO13969 |
| 16 | 37.6 | 3.5 | 6310  | 24 | AB170322 |
| 17 | 37.6 | 3.5 | 6310  | 24 | AA561269 |
| 18 | 37.6 | 3.5 | 6310  | 24 | ABK13365 |
| 19 | 36.6 | 3.4 | 6151  | 23 | AB106707 |
| 20 | 36.6 | 3.4 | 9173  | 23 | AB106706 |
| 21 | 36.2 | 3.3 | 5059  | 20 | AA84332  |
| 22 | 36   | 3.3 | 10732 | 21 | AA10594  |
| 23 | 35.6 | 3.3 | 904   | 16 | AA54849  |
| 24 | 35.2 | 3.2 | 640   | 24 | ABO37720 |
| 25 | 35.2 | 3.2 | 640   | 24 | ABO37721 |
| 26 | 35.2 | 3.2 | 927   | 22 | AAH50774 |
| 27 | 35.2 | 3.2 | 951   | 22 | AAH50883 |
| 28 | 35.2 | 3.2 | 2814  | 21 | AAZ50960 |
| 29 | 35.2 | 3.2 | 19718 | 19 | AAV5232  |
| 30 | 35   | 3.2 | 864   | 23 | AA593326 |
| 31 | 35   | 3.2 | 2257  | 23 | AA589627 |
| 32 | 35   | 3.2 | 2546  | 23 | AA590013 |
| 33 | 34.6 | 3.2 | 475   | 20 | AAV88421 |
| 34 | 34.6 | 3.2 | 50341 | 19 | AAV2674  |
| 35 | 34.6 | 3.2 | 50341 | 21 | AAZ39519 |
| 36 | 34.6 | 3.2 | 52297 | 16 | AAQ47357 |
| 37 | 34.6 | 3.2 | 52298 | 14 | AAQ47357 |
| 38 | 34.6 | 3.2 | 52298 | 14 | AAQ47357 |
| 39 | 34   | 3.1 | 2855  | 22 | AA531515 |
| 40 | 34   | 3.1 | 2855  | 24 | ABO66839 |
| 41 | 33.4 | 3.1 | 775   | 24 | ABN81171 |
| 42 | 33.4 | 3.1 | 2895  | 24 | AA590365 |
| 43 | 33.4 | 3.0 | 2895  | 23 | AA593337 |
| 44 | 33.2 | 3.0 | 498   | 24 | ABN781   |
| 45 | 33.2 | 3.0 | 601   | 23 | AB109003 |

#### ALIGNMENTS

|          |   |
|----------|---|
| RESULT 1 |   |
| AAT33504 | standard; DNA: 13473 BP.  |
| 16       | AAT33504;   |
| 17       | 18-NOV-1996 (first entry)   |
| 18       | Infectious laryngotracheitis virus short region sequence.         |
| 19       | Infectious laryngotracheitis virus; ILTV; herpesvirus;            |
| 20       | attenuation; vector; vaccine; chicken; poultry; immunisation; ds. |
| 21       | Infectious laryngotracheitis virus.                               |
| 22       | Key   |
| 23       | repeat_region   |
| 24       | Location/Qualifiers   |
| 25       | 1..273  |
| 26       | /*tag- a  |
| 27       | /*function= internal repeat region                                |
| 28       | 274..13371  |
| 29       | /*tag- b  |
| 30       | /*function= unique short sequence of ILTV                         |
| 31       | complement (281..970)   |
| 32       | /*tag- c  |
| 33       | /*label= US2-gene   |
| 34       | 1059..2489  |
| 35       | /*tag- d  |
| 36       | /*label= Protein-kinase_gene                                      |
| 37       | 2575..4107  |
| 38       | /*tag- e  |
| 39       | /*label= UL47-1like_gene  |
| 40       | 4113..4445  |
| 41       | CDS   |
| 42       | Drosophila melanog  |
| 43       | Drosophila melanog  |
| 44       | Drosophila melanog  |
| 45       | Drosophila melanog  |



|   |    |  |
|---|----|--|
|   | FT | /tag= f.   |
|   | FT | /label= ORF4   |
| CDS   | FT | complement (4139..4519)  |
|   | FT | /tag= g  |
|   | FT | /label= ORF4_reverse_complement                                    |
| misc_difference   | FT | 4535   |
|   | FT | *tag= h  |
|   | FT | /note= "base 4535 is given as 's' in the specification"            |
| CDS   | FT | 4609..5487   |
|   | FT | /tag= i  |
|   | FT | /label= gg_gene  |
| CDS   | FT | 5697..8654   |
|   | FT | *tag= j  |
|   | FT | /label= g60_gene   |
| CDS   | FT | complement (6948..7826)  |
|   | FT | /tag= k  |
|   | FT | /label= ORF6_reverse_complement                                    |
|   | FT | 8462..9766   |
|   | FT | *tag= l  |
|   | FT | /label= gd_gene  |
| CDS   | FT | 9874..10962  |
|   | FT | *tag= m  |
|   | FT | /label= gi_gene  |
| CDS   | FT | complement (10617..11150)  |
|   | FT | *tag= n  |
|   | FT | /label= ORF8_reverse_complement                                    |
| CDS   | FT | 11159..12658   |
|   | FT | *tag= o  |
|   | FT | /label= ge_gene  |
| CDS   | FT | 12665..13447   |
|   | FT | /tag= p  |
|   | FT | /label= ORF10  |
| misc_difference   | FT | 13002..13003   |
|   | FT | *tag= q  |
|   | FT | /note= "bases 13002-13303 are given as 'ss' in the specification"  |
| repeat_region   | FT | 13372..13473   |
|   | FT | *tag= r  |
|   | FT | /function= terminal repeat region                                  |
| W09508622-A1.   | XX |  |
| 30-MAR-1995.  | PD |  |
| 16-SEP-1994:  | PF | 94MO-US10628.  |
| 24-SEP-1993:  | PR | 93US-O126597.  |
| (SVTR ) SYNPRO CORP.  | XX |  |
| Cochran MD, Wild MA:  | PI |  |
| WI; 1995-139591/18.   | DR |  |
| p-PsDB; AAM00630, AAM00631, AAM00632, AAM00633, AAM00634, AAM00635, | DR | AAM00636, AAM00637, AAM00638, AAM00639, AAM00640, AAM00641 W00642. |
| Recombinant attenuated infectious lagynotracheitis virus - for use  | PT |  |
| in vaccines to protect poultry from infection from the virus, also  | PT |  |
| methods of distinguishing between vaccinated and naturally infected | PT |  |
| birds   | PT |  |
| Example 1: Page 79-94; 177pp: English.                              | XX |  |
| The unique short region (AAT33504) of infectious lagynotracheitis   | CC |  |
| virus (ILTV) genomic DNA contains genes (see also AAT33505 and      | CC |  |
| AAT33510-13) that are associated with ILTV virulence. A deletion in | CC |  |
| those genes, esp. the glycoprotein g6 gene, glycoprotein g1 gene,   | CC |  |
| thymidine kinase gene, US2 gene, UL47-like gene or the glycoprotein | CC |  |
| g60 gene, will attenuate the ILTV. A gene for a foreign antigen may | CC |  |
| be inserted into the US2, UL47-like, ORF4, gc, g60 or g1 gene to    | CC |  |
| produce a recombinant ILTV. Recombinant ILTV may be used as a       | CC |  |
| multivalent vaccine, esp. for use in poultry.                       | CC |  |

| xx | Sequence   | 134/73 BP; 3390 A; 3582 C; 3547 G; 2951 T; 3 other: |
|----|--|---|
|    | Query Match  | 100.0%; Score 1089; DB 16; Length 134/73;           |
|    | Best Local Similarity  | 100.0%; Pred. No. 0;                                |
|    | Matches 1089; Conservative   | 0; Mismatches 0; Indels 0; Caps 0;                  |
| OY | 1 ATGGCATTCCGCTACTTGGAGACTCTCTGGCTCTCTTGCCCGGACAGCTCGGACACCTTCGGGGCG | 60  |
| Db | 9874 ATGGCATCGCTACTTGGAACTCTGGCTCTCTTCCGCGAGCGCTCGACCTTCGGGGCG       | 9933  |
| OY | 61 ATGGGAATTCGATACACTGTGAAATCAGTCTCCGCCAGATTGACGACGATCATCATCTGT      | 120   |
| Db | 9934 ATGGGAATTCGATACACTGTGAAATCAGTCTCCGCCAGATTGACGACGATCATCATCTGT    | 9993  |
| OY | 121 ATCGTCGGCTCGCCCCGAAAGCTACAAATTCAGTCACTATTTTTCATGCTGGCCAG         | 180   |
| Db | 9994 ATCTCCGCGCTCCCGCGAAGCTACAAATTCAGTCACTATTTTTCATGCGTGGCCAG        | 10053   |
| OY | 181 AGACCCCAAAACCTTACTGAGAAACGTCGCGGCTCGCTTGGGTGTGATATTAACAAC        | 240   |
| Db | 10054 AGACCCCAAAACCTTACTGAGAAACGTCGCGGCTCGCTTGGGTGTGATATTAACAAC      | 10111   |
| OY | 241 CAGTGTACAGAGAACTTACGAGAGACGCTTTGAAAAATTGCACTCATGACTGTCTCT        | 300   |
| Db | 10114 CAGTGTACAGAGAACTTACGAGAGACGCTTTGAAAAATTGCACTCATGACTGTCTCT      | 10177   |
| OY | 301 GTTTTGTGGGCTGTAAAGTGACCGAGTACACGTTCCGCTCGAACAACATTAACCGGA        | 360   |
| Db | 10174 GTTTTGTGGGCTGTAAAGTGACCGAGTACACGTTCCGCTCGAACAACATTAACCGGA      | 10233   |
| OY | 361 CCTCCACACCCGTTTAAAGCTCACTATACGAATCTCTGCCGAACGACACGGGATGTT        | 420   |
| Db | 10234 CCTCCACACCCGTTTAAAGCTCACTATACGAATCTCTGCCGAACGACACGGGATGTT      | 10293   |
| OY | 421 TAGCTAATTTGTGGGCTGAGACGACACCAAAAGAACCTTACGCTTCGGGATCCAACTA       | 480   |
| Db | 10294 TAGCTAATTTGTGGGCTGAGACGACACCAAAAGAACCTTACGCTTCGGGATCCAACTA     | 10353   |
| OY | 481 TCGGTGTATCAATTCGCGAACACCGCCGCGACTCGCGACTTATTTCCAAAGGCTGTGT       | 540   |
| Db | 10354 TCGGTGTATCAATTCGCGAACACCGCCGCGACTCGCGACTTATTTCCAAAGGCTGTGT     | 10411   |
| OY | 541 CGCACCTTCGGATTACTACCGTCCAACTTGAGGCTATCTCAGGACCGAAGAAATGTG        | 600   |
| Db | 10414 CGCACCTTCGGATTACTACCGTCCAACTTGAGGCTATCTCAGGACCGAAGAAATGTG      | 10477   |
| OY | 601 CGCAACTGGCAAGCGTACGTTGCCACGAGAGCCACGACGACACGCGCCGAGGAGACAAC      | 660   |
| Db | 10474 CGCAACTGGCAAGCGTACGTTGCCACGAGAGCCACGACGACACGCGCCGAGGAGACAAC    | 10533   |
| OY | 661 CCGACGCCCGTCACTGCACACGAGCGCTCCGAACCTTGAGCGGAACTTTACCTTTCC        | 720   |
| Db | 10534 CCGACGCCCGTCACTGCACACGAGCGCTCCGAACCTTGAGCGGAACTTTACCTTTCC      | 10593   |
| OY | 721 TGGGTAGAAAATGGCGGATCATTACGAACCGACACCCGCAAAAGAAAATTCAAAAGTT       | 780   |
| Db | 10594 TGGGTAGAAAATGGCGGATCATTACGAACCGACACCCGCAAAAGAAAATTCAAAAGTT     | 10655   |
| OY | 781 ACTGTCCGTCTCGGGAATGAGCCCTACGCTAATTGGGTTAACCGTGGCTGCCGTG          | 840   |
| Db | 10654 ACTGTCCGTCTCGGGAATGAGCCCTACGCTAATTGGGTTAACCGGCTGCCGTG          | 10711   |
| OY | 841 AGCCCAACGATCGGCGCTGCTCATTTGTAATTTCCATGTCACGAAAACATGTCCACCCG      | 900   |
| Db | 10714 AGCCCAACGATCGGCGCTGCTCATTTGTAATTTCCATGTCACGAAAACATGTCCACCCG    | 10777   |
| OY | 901 CACGGAATAATTAGACAGGTCTCGCAAGACGACGAAGAAGTTCCCAAATTAAAGGAA        | 960   |
| Db | 10774 CACGGAATAATTAGACAGGTCTCGCAAGACGACGAAGAAGTTCCCAAATTAAAGGAA      | 10833   |
| OY | 961 TCGGGAATAATTGGACCAATGTTGCTGCGAAATTAACAAGGGCGCTGCACAGGATAGT       | 1020  |

```

DB      10834 TCGGGAAATTTGGACCCGATGTTGCTGCGAATAAACAAGGCGCTGACCAAGATAGT 10893
OY      1021 GAACCTTGAGAACCTGCTGGATTTGTTAACCCGCTGACGCTAAGCTCCGCACTCAATA 10880
DB      10894 GAACCTTGAGAACCTGCTGGATTTGTTAACCCGCTGACGCTAAGCTCCGCACTCAATA 10953
OY      1081 AAAATGTGA 1089
DB      10954 AAAATGTGA 10962

RESULT 2
AAAT4384
ID      AAAT4384 standard; DNA; 13473 BP.
XX      AAAT4384;
XX      02-JUN-1997 (first entry)
XX      Infectious laryngotracheitis virus unique short region.
XX      ILTV; vaccine; vector; attenuation; poultry;
XX      avian infectious bronchitis virus; Newcastle disease virus;
XX      infectious bursal disease virus of chickens;
XX      Marek's disease virus; herpesvirus; ss.
XX      Infectious laryngotracheitis virus USDA strain 8302.
XX      Key
XX      Location/Qualifiers
XX      repeat_region
XX      1..273
XX      /tag= a
XX      274..13371
XX      /tag= b
XX      /note= "unique short region"
XX      151..166
XX      /tag= c
XX      /note= "polyA signal for ORF1"
XX      complement (281..970)
XX      CDS
XX      /tag= d
XX      /label= ORF1(RC)
XX      /note= "unique-like short 2 (US2) gene"
XX      1007..1010
XX      /tag= e
XX      /note= "TATA signal for ORF1 and ORF2"
XX      1040..1043
XX      /tag= f
XX      /note= "TATA signal for ORF1"
XX      1042..1045
XX      /tag= g
XX      /note= "TATA signal for ORF2"
XX      1059..2489
XX      /tag= h
XX      /label= ORF2
XX      /note= "protein kinase gene"
XX      2491..2496
XX      /tag= i
XX      /note= "polyA signal for ORF2"
XX      2538..2543
XX      /tag= j
XX      /note= "polyA signal for ORF2"
XX      2575..4107
XX      /tag= k
XX      /label= ORF3
XX      /note= "unique long 47 (UL47)-like gene"
XX      4523..4526
XX      /tag= l
XX      /note= "TATA signal for ORF4"
XX      4113..4445
XX      /tag= m
XX      /label= ORF4
XX      complement (4139..4519)
XX      /tag= n
XX      /label= ORF4(RC)
XX      CDS
XX      /label= ORF4(RC)
XX      26-SEP-1996.

FT      TATA_signal
FT      4523..4526
FT      /tag= o
FT      /note= "TATA signal for ORF5"
FT      4609..5487
FT      /tag= p
FT      /label= ORF5
FT      /note= "glycoprotein g6 gene"
FT      4609..4686
FT      /tag= q
FT      4687..5484
FT      /tag= r
FT      5564..5569
FT      /tag= s
FT      /note= "polyA site for ORF5"
FT      5653..5658
FT      /tag= t
FT      /note= "polyA signal for ORF5"
FT      5697..8654
FT      /tag= u
FT      /label= ORF6
FT      /note= "glycoprotein g60 gene"
FT      6987..7727
FT      /tag= v
FT      /note= "repeat region consists of approx. 23
FT      repeats of 30-36 bp"
FT      complement (6948..7826)
FT      /tag= w
FT      /label= ORF6(RC)
FT      8455..8458
FT      /tag= x
FT      /note= "potential TATA signal for ORF7"
FT      8461..9766
FT      /tag= y
FT      /label= ORF7
FT      /note= "glycoprotein gD gene, alternative start
FT      9819..9822
FT      /tag= z
FT      /note= "TATA signal for ORF8"
FT      9874..10962
FT      /tag= aa
FT      /label= ORF8
FT      /note= "glycoprotein gI gene"
FT      9874..9939
FT      /tag= ab
FT      9940..10959
FT      /tag= ac
FT      complement (10617..11150)
FT      /tag= ad
FT      /label= ORF8(RC)
FT      11069..11072
FT      /tag= ae
FT      /note= "TATA signal for ORF9"
FT      11159..12658
FT      /tag= af
FT      /label= ORF9
FT      /note= "glycoprotein gE gene"
FT      11159..11212
FT      /tag= ag
FT      11213..12655
FT      /tag= ah
FT      12483..12486
FT      /tag= ai
FT      /note= "TATA signal for ORF10"
FT      12865..13447
FT      /tag= aj
FT      /label= ORF10
FT      13372..13743
FT      /tag= ak
FT      26-SEP-1996.

```



```

FT      repeat_region      1..2909
FT      /tag- a
FT      /note- "short repeat region"
FT      697..1533
FT      /tag- b
FT      /label- SRORF2
FT      /note- "US10 gene"
FT      complement (2916..3605)
FT      CDS
FT      /tag- c
FT      /label- SRORF1
FT      /note- "short repeat open reading frame 1"
FT      2910..16003
FT      /tag- d
FT      /note- "unique short region"
FT      2786..2881
FT      /tag- e
FT      /note- "polyA signal for ORF1"
FT      complement (2916..3605)
FT      CDS
FT      /tag- f
FT      /label- ORF1(RC)
FT      /note- "unique-like short 2 (US2) gene"
FT      3642..3645
FT      /tag- g
FT      /note- "TATA signal for ORP1 and ORP2"
FT      3675..3678
FT      /tag- h
FT      /note- "TATA signal for ORP1"
FT      3677..3680
FT      /tag- i
FT      /note- "TATA signal for ORP2"
FT      3694..5124
FT      /tag- j
FT      /label- ORF2
FT      /note- "protein kinase gene"
FT      5126..5131
FT      /tag- k
FT      /note- "polyA signal for ORF2"
FT      5173..5178
FT      /tag- l
FT      /note- "polyA signal for ORF2"
FT      5210..7081
FT      CDS
FT      /tag- m
FT      /label- ORF3
FT      /note- "unique long 47 (UL47)-like gene"
FT      7158..7161
FT      /tag- n
FT      /note- "TATA signal for ORP4"
FT      6748..7080
FT      /tag- o
FT      /label- ORF4
FT      /note- complement (6774..7154)
FT      CDS
FT      /tag- p
FT      /label- ORF4(RC)
FT      7158..7161
FT      /tag- q
FT      /note- "TATA signal for ORF5"
FT      7245..8123
FT      /tag- r
FT      /label- ORF5
FT      /note- "glycoprotein gg gene"
FT      7245..7332
FT      /tag- s
FT      7333..8120
FT      /tag- t
FT      8200..8205
FT      /tag- u
FT      /note- "polyA site for ORF5"
FT      8289..8294
FT      /tag- v
FT      /note- "polyA signal for ORF5"
FT      8333..11290
FT      /tag- w
FT      /label- ORF6

FT      repeat_region      /note- "glycoprotein g60 gene"
FT      9623..10363
FT      /tag- x
FT      /note- "repeat region consists of approx. 23
FT      repeats of 30-36 bp"
FT      complement (9584..10462)
FT      CDS
FT      /tag- y
FT      /label- ORF6(RC)
FT      11091..11094
FT      /tag- z
FT      /note- "potential TATA signal for ORF7"
FT      11098..12402
FT      /tag- aa
FT      /label- ORF7
FT      12455..12448
FT      /tag- ab
FT      /note- "TATA signal for ORF8"
FT      12510..13598
FT      /tag- ac
FT      /label- ORF8
FT      /note- "glycoprotein gi gene"
FT      12510..12575
FT      /tag- ad
FT      12576..13595
FT      /tag- ae
FT      complement (13253..13786)
FT      /tag- af
FT      /label- ORF8(RC)
FT      13705..13708
FT      /tag- ag
FT      /note- "TATA signal for ORF9"
FT      13792..15291
FT      /tag- ah
FT      /label- ORF9
FT      /note- "glycoprotein ge gene"
FT      13792..13845
FT      /tag- ai
FT      13846..15288
FT      /tag- aj
FT      15116..15119
FT      /tag- ak
FT      /note- "TATA signal for ORF10"
FT      15298..16080
FT      /tag- al
FT      /label- ORF10
FT      16004..18912
FT      /tag- am
FT      16129..17013
FT      /tag- an
FT      complement (17380..18216)
FT      /tag- ao

FT      repeat_region      W096293396-A1.
FT      26-SEP-1996.
FT      21-MAR-1996; 96WO-US03916.
FT      06-JUN-1995; 95US-0468190.
FT      23-MAR-1995; 95US-0410121.
FT      (SYTR ) SYNPRO CORP.
FT      Cochran MD, Wild MA.
FT      WPI: 1996-443372/44.
FT      P-FSD: AAM06782, AAM06783, AAM06784, AAM06785, AAM06786, AAM06787,
FT      AAM06788, AAM06789, AAM06790, AAM06791, AAM06792, AAM06793, AAM01415,
FT      AAM01416.
FT      Recombinant infectious laryngotracheitis virus with deletion in the
FT      glycoprotein G, gi or US2 gene, etc. - useful for vaccines against
FT      infectious laryngotracheitis in poultry

```

Example 11; Page 138-154; 216pp; Eng11sh.

CC The nucleotide sequence of 19,912 bp of contiguous DNA (AA144385)  
CC from the unique short and flanking region of infectious  
CC largeorthorectus virus (ILTV) contains the entire 13,098 bp unique  
CC short region (see also AA144384) and includes 17 open reading frame  
CC that encode proteins (AAW06782-94, AAW0115-16) of over 100 amino  
CC acids, 10 of which show homology to other virus genes. Novel  
CC recombinant, attenuated ILTV comprises the ILTV genome cont. a  
CC deletion in the unique short region, esp. in the glycoprotein g5,  
CC g1, US2, ORR4, U47-1like or g60 gene. The attenuated virus is  
CC useful as a vaccine against ILTV. A foreign gene encoding an  
CC antigen e.g. from another avian virus can be inserted into the US2,  
CC U47-1like, ORR4, g5, g60 or g1 gene to provide a multivalent  
CC vaccine for chickens and other poultry. Deletion of the g5 or g1  
CC gene provides as a negative marker to distinguish vaccinated from  
CC infected animals.

Sequence 18912 BP; 4416 A; 5261 C; 5251 G; 3984 T; 0 other;

|                       |              |              |            |              |
|-----------------------|--------------|--------------|------------|--------------|
| Query Match           | 99.9%        | Score 1087.4 | DB 17      | Length 18912 |
| Best local similarity | 99.9%        | Pred. No. 0  |            |              |
| Matches 1088          | Conservative | 0            | Mismatches | 1            |
|                       |              |              | Indels     | 0            |
|                       |              |              | Gaps       | 0            |

|    |       |   |       |
|----|-------|---|-------|
| QY | 1     | ATGGCAATCGCTACTTGGAACTCGTGGCTCCCTTGGCCGACGCTGCACCTTTCGGCGGG     | 60    |
| Db | 12510 | ATGGCAATCGCTACTTGGAACTCGTGGCTCCCTTGGCCGACGCTGCACCTTTCGGCGGG     | 12566 |
| QY | 61    | ATGGGAATCGTGTATCACTGGAAATCAACGCTCCGCCAGGATTTGACGACATCATCTGTG    | 120   |
| Db | 12570 | ATGGGAATCGTGTATCACTGGAAATCAACGCTCCGCCAGGATTTGACGACATCATCTGTG    | 12629 |
| QY | 121   | ATGCTGCGCGCTCGGCCCGGAAGCTACAAATTCAACCTGACGTATTTTCAATGGCTGGCCAG  | 180   |
| Db | 12630 | ATGCTGCGCGCTCGGCCCGGAAGCTACAAATTCAACCTGACGTATTTTCAATGGCTGGCCAG  | 12688 |
| QY | 181   | AGACCCCAACAAACCCCTACTCAGAACCGTCCGGCTGCGCTTTCGCTGATATTAACAAAC    | 240   |
| Db | 12690 | AGACCCCAACAAACCCCTACTCAGAACCGTCCGGCTGCGCTTTCGCTGATATTAACAAAC    | 12749 |
| QY | 241   | CAATGCTACCGAGACACTTGAACGAGAGCGCTTGAATAATTGCACTCATATGATCGCTTCT   | 300   |
| Db | 12750 | CAATGCTACCGAGACACTTGAACGAGAGCGCTTGAATAATTGCACTCATATGATCGCTTCT   | 12809 |
| QY | 301   | GTTTTTGTGCGCTGTAAAGTGACCGAGTACACGTTCTCCGCTCTGAAACAGACTTAACCGGA  | 360   |
| Db | 12810 | GTTTTTGTGCGCTGTAAAGTGACCGAGTACACGTTCTCCGCTCTGAAACAGACTTAACCGGA  | 12869 |
| QY | 361   | CTTCCACACCCCGTTTAAAGTCACTATACGAATTCCTCTGTCGGAACGACACGCGGATGTTTC | 420   |
| Db | 12870 | CTTCCACACCCCGTTTAAAGTCACTATACGAATTCCTCTGTCGGAACGACACGCGGATGTTTC | 12929 |
| QY | 421   | TACGTAATTTGTTGGCTAGACGACACCAAGAACCCATTGAGCTCTTCGCAATCCAACTA     | 480   |
| Db | 12930 | TACGTAATTTGTTGGCTAGACGACACCAAGAACCCATTGAGCTCTTCGCAATCCAACTA     | 12989 |
| QY | 481   | TGCGTGTATCAATTCCGGAACACCCCGCGACTCGCGGACTTATTTCCAAAGCTTCGTGT     | 540   |
| Db | 12990 | TGCGTGTATCAATTCCGGAACACCCCGCGACTCGCGGACTTATTTCCAAAGCTTCGTGT     | 13049 |
| QY | 541   | CGACACTTGGGATTACTCTACCGTCCAACTTGAAGGCTATCTCAGGACCGAGGAAGTTGG    | 600   |
| Db | 13050 | CGACACTTGGGATTACTCTACCGTCCAACTTGAAGGCTATCTCAGGACCGAGGAAGTTGG    | 13109 |
| QY | 601   | CGCAACTGGCAAGCGTACGTTGGCAACGAGGCGACAGACGACGACGCGAGCGACAAC       | 660   |
| Db | 13110 | CGCAACTGGCAAGCGTACGTTGGCAACGAGGCGACAGACGACGACGCGAGCGACAAC       | 13169 |
| QY | 661   | CGAGGCGCGCTACTACCTGCAACGAGCGCTCCGAATTTGAAGCGGAACACTTTACCTTTCC   | 720   |
| Db | 13170 | CGAGGCGCGCTACTACCTGCAACGAGCGCTCCGAATTTGAAGCGGAACACTTTACCTTTCC   | 13229 |

|    |       |   |       |
|----|-------|---|-------|
| QY | 721   | TGGGTAGAAATGGCGTGGATTCATTAGAACCGACACCCGCCAAGAAAATTCAAACGTT    | 780   |
| Db | 13230 | TGGGTAGAAATGGCGTGGATTCATTAGAACCGACACCCGCCAAGAAAATTCAAACGTT    | 13289 |
| QY | 781   | ACTGTCGCTCGGGACAAATGAGACCCCTACGCTAATTGGGGTAACCTGTGCTGCCGTCGTG | 840   |
| Db | 13290 | ACTGTCGCTCGGGACAAATGAGACCCCTACGCTAATTGGGGTAACCTGTGCTGCCGTCGTG | 13349 |
| QY | 841   | AGCCCAACGATCGGCGCTCTGTCATTTGTAAATTCATGTCACCAAGAAATGTGACACCCG  | 900   |
| Db | 13350 | AGCCCAACGATCGGCGCTCTGTCATTTGTAAATTCATGTCACCAAGAAATGTGACACCCG  | 13409 |
| QY | 901   | CACGGAATTTAGACACGTCCTGCGCAAGACGACGAAGAAACGTTCCCAACTGGAAGGAA   | 960   |
| Db | 13410 | CACGGAATTTAGACACGTCCTGCGCAAGACGACGAAGAAACGTTCCCAACTGGAAGGAA   | 13469 |
| QY | 961   | TCGCGAAATTTGGACCCATGTTGTCGTGCGAATAAACAAAGGGCGCTGACCAGATAGT    | 1020  |
| Db | 13470 | TCGCGAAATTTGGACCCATGTTGTCGTGCGAATAAACAAAGGGCGCTGACCAGATAGT    | 13529 |
| QY | 1021  | GAACCTTGGAACCTGCTTGCATTTGTTAACCCGCTTGCCTTAAGCTGCCCGACCTCATTA  | 1080  |
| Db | 13530 | GAACCTTGGAACCTGCTTGCATTTGTTAACCCGCTTGCCTTAAGCTGCCCGACCTCATTA  | 13589 |
| QY | 1081  | AAATGTGA 1089   |       |
| Db | 13590 | AAATGTGA 13598  |       |

#### RESULT 4

ID AAT33513 standard; DNA; 534 BP.

AC AAT33513;

DE Infectious laryngotracheitis virus ORF8 reverse complement.

KW Infectious laryngotracheitis virus; ILTV; herpesvirus;

KW  
ds.  
KW

OS Infectious laryngotracheitis virus.

PN W09508622-A1

PD 30-MAR-1995.

PF 16-SEP-1994; 94WO-US10628.

PR 24-SEP-1993; 93US-0126597.

PA (SYTR ) SYNTRO CORP.  
VV

PI Cochran MD, Wild MA;

DR WPI; 1995-139591/18

**XX**

PT Recombinant attenuated infectious laryngotracheitis virus - for use  
PT in vaccines to protect poultry from infection from the virus, also  
PT methods of distinguishing between vaccinated and naturally infected  
PT birds

PS Example 1; Page 115-116; 177pp; English

CC ORR8 reverse complement (AAT33513) consists of bases 11,150 to  
CC 10,617 on the reverse complement strand of the unique short  
CC region (see also AAT33504) of infectious laryngotracheitis virus  
CC (ILTV) genomic DNA. It codes for a protein (AAM00642) of  
CC approx. 19,470 mol.wt. of unknown function.



SO Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
Query Match 3.5%; Score 38.2; DB 22; Length 4411529;  
Best Local Similarity 58.3%; Pred. No. 8.9;  
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
OY 29 TCCTTCGCGGAGCGCTGCGACCTTCGCGCGATGGGAATCGTATCAGTGAATCAGC 88  
DB 618034 TCCGACGCGCGACGCTGCGACCTTCGCGCGATGGGAATCAGTGAATCAGC 88  
OY 89 TCTCCGCGAGATTGACGAGATCAATCGTATCGTGGCGCTCCGCGCGAAGC 143  
DB 617974 TCGGCCCCAGGACGAGTGTCTCAGCCGCGGAAACACCGCAGCAGCGGCGC 617920

RESULT 7  
AAF26455  
ID AAF26455 standard; cDNA; 2526 BP.  
XX AAF26455;  
XX 26-MAR-2001 (first entry)  
XX D. melanogaster corkscrew DNA.  
DE SHP-2; SHP-1; Src Homology-2; protein tyrosine phosphatase; mutant;  
KW neoplastic disorder; obesity; angiogenesis; cancer; immune;  
KM hematopoietic; allergy; ss.  
XX Drosophila melanogaster.  
OS US6156551-A.  
PN 05-DEC-2000.  
XX 05-JUN-1998; 98US-0092443.  
XX 05-JUN-1998; 98US-0092443.  
PR 05-JUN-1998; 98US-0092443.  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
PA (JOSL-) JOSLIN DIABETES CENT.  
XX Neel BG, Shoelson S, Pluskey S, O'Reilly AM;  
XX WPI; 2001-060166/07.  
DR Mutant SH2 domain-containing protein tyrosine phosphatase, useful in in  
XX vitro assays to screen for binding partners, inhibitors of tyrosine  
PT phosphatase and for treating tyrosine phosphatase-mediated diseases -  
XX Disclosure; Fig 7; 161pp; English.  
XX The present invention relates to an activated SH2 (Src Homology-2)  
CC -domain containing protein tyrosine phosphatase (SHP-1 or SHP-2)  
CC mutant with a mutation in the SH2 domain. Activated mutants of  
CC SH2-domain-containing protein tyrosine phosphatases are useful in  
CC in vitro assays to screen for binding partners and inhibitors of  
CC the phosphatase and in the treatment of PPP-mediated diseases or  
CC conditions in a mammal, including neoplastic disorders, obesity and  
CC to inhibit angiogenesis. Inhibitors identified using the activated  
CC mutants are useful for the treatment of cancer, immunosuppression,  
CC immunostimulation, hematopoietic stimulation and anti-allergy  
CC treatment.  
XX Sequence 2526 BP; 616 A; 742 C; 732 G; 436 T; 0 other;  
SO

Query Match 3.5%; Score 38; DB 22; Length 2526;  
Best Local Similarity 51.8%; Pred. No. 0.26;  
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
OY 569 TTGAGGCTTATCTCAGAGCCGAGGAAAGTTGGCGCAACTGCGTACGTTGCCACGG 628  
DB 1325 TTAAAGACTTACATCGCACCCAGGCGTGTCTGCTCACCCAGCAAGTGAACACGCGTACGG 1384

OY 629 AGGCCACGACGACCGCCGAGCGACGAAACCCGACGCCGTACTGCAACCGAGCGCT 688  
DB 1385 ACTTTCGGAACATGCTGTGCGACGAGAAACGCGGGGTGATGTCATGACCAAGAGT 1444  
OY 689 CCGAAGCTTGAAGCGGAACACTTACCTTCCCTGCGTAGAAATGC 734  
DB 1445 ACGAGCGCGCGCAAAAGAAAGTGGCCCGCTACTGCGCGGACGAGGG 1490

RESULT 8  
ABL06927  
ID ABL06927 standard; cDNA; 2677 BP.  
XX ABL06927;  
XX 26-MAR-2002 (first entry)  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 15263.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
PN 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EM;  
XX WPI; 2001-656860/75.  
DR P-P-SDB; ABB62824.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX Claim 1; SEQ ID NO 15263; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU130511), expressed DNA  
CC sequences (ABU1840-ABU16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 2677 BP; 670 A; 760 C; 749 G; 498 T; 0 other;  
SO

Query Match 3.5%; Score 38; DB 23; Length 2677;  
Best Local Similarity 51.8%; Pred. No. 0.27;  
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
OY 569 TTGAGGCTTATCTCAGAGCCGAGGAAAGTTGGCGCAACTGCGTACGTTGCCACGG 628  
DB 1464 TTAAAGACTTACATCGCACCCAGGCGTGTCTGCTCACCCAGCAAGTGAACACGCGTACGG 1523  
OY 629 AGGCCACGACGACCGCCGAGCGGACGAAACCCGAGCGCGCTGATGTCGACCAAGAGCT 688  
DB 1524 ACTTTCGGAACATGCTGTGCGACGAGAAACGCGGGGTGATGTCATGACCAAGAGT 1583







XX 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, LI PWD, Myers EW;  
XX WPI: 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 14875; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB16175-AB16175).  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 21399 BP; 6083 A; 5128 C; 4585 G; 5603 T; 0 other;  
SQ  
Query Match 3.5%; Score 38; DB 23; Length 21399;  
Best Local Similarity 51.8%; Pred. No. 0.75; Mismatches 86; Conservative 0; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
XX  
XX 569 TTGAGCCTATCTCAGACGAGGAAAGTTGGCGCAACTGCAACCTGCTGCGACGG 628  
DB 16998 TTAAGACCTACATCCGCCACCGGCTGTCTGTCACCCGCAAGTGAACACGGTGACGG 17057  
XX  
XX 629 AGGCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 688  
DB 17058 ACTTCTGGGAACATGCTGCGGAGGAAACGCGGGGTGATCTCATGTACACGACGAGAGT 17117  
XX  
XX 689 CCGAACTGAAAGCGGAAACATTTACTCTTCCCTGCTAGAAATG 734  
DB 17118 ACGAGCGCGGCAAGAAAGTGGCCGCTACTGCGGCGGACGAGGG 17163  
XX  
XX  
XX RESULT 14  
XX ABQ13968/C  
XX ID ABQ13968 standard; DNA; 813 BP.  
XX  
XX ABQ13968;  
XX  
XX 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 559.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX  
XX WO200218632-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001MO-EP10074.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX

PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Plepenbrock C, Berlin K, Guettig D;  
XX WPI: 2002-371829/40.  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP s); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ34121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX Sequence 813 BP; 114 A; 99 C; 350 G; 250 T; 0 other;  
SQ  
Query Match 3.5%; Score 37.6; DB 24; Length 813;  
Best Local Similarity 49.0%; Pred. No. 0.2; Mismatches 100; Conservative 0; Mismatches 104; Indels 0; Gaps 0;  
Matches 100; Conservative 0; Mismatches 104; Indels 0; Gaps 0;  
XX  
XX 611 AAGGCTAGCTTGGCAGGAGGACGACGACGACGACGACGACGACGACGACGACGACG 670  
DB 772 AAAGCCCCCGGACCCGCAATACAAACGACGACGACGACGACGACGACGACGACGACG 713  
XX  
XX 671 TCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 730  
DB 712 ACCGCGATCCCGGACCCGCAATACAAACGACGACGACGACGACGACGACGACGACG 653  
XX  
XX 731 ATGCGGTGATCATTTACGAAACGACGACGACGACGACGACGACGACGACGACGACG 790  
DB 652 ATACTACACCGACTACGAAACGACGACGACGACGACGACGACGACGACGACGACG 790  
XX  
XX 791 TCGGACAAATGAGCCCTAGGCTTA 814  
DB 592 GCCGCAAAAAGCCCTCCGACAA 599  
XX  
XX  
XX RESULT 15  
XX ABQ13969  
XX ID ABQ13969 standard; DNA; 813 BP.  
XX  
XX ABQ13969;  
XX  
XX 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 560.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.  
YY

|    |                              |
|----|------------------------------|
| PR | 01-SEP-2000; 2000DE-1043826. |
| PR | 05-SEP-2000; 2000DE-1044543. |
| PR | 05-SEP-2000; 2000DE-1044543. |

XX (EPITG-) EPITGENOMICS AG  
PA

|    |      |           |
|----|------|-----------|
| XX | 0106 | Blanchard |
| PT | 0107 | Blanchard |

XX WPT: 2002-271020 / 00  
DB

**XX**      **Determining the degree**

amplicons from chemically treated DNA -

Claim 12; 56pp + Sequence Listing; 56pp; German

CC This invention describes a novel method for determining the degree of contribution of a particular attribute to a model.

methylation. Of a particular, cytosine in a molting -Cp-3' present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (1) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations of single nucleotide polymorphisms (SNPs); and (11) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABO1310-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

**SQ** Sequence 813 BP; 250 A; 350 C; 99 G; 114 T; 0 other;

| Query Match | Score | DB  | Length |
|-------------|-------|-----|--------|
| 3.58;       | 37.6; | 24; | 813;   |

```
Matches 100; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
```

611 AAGCGTACGTTGCCACGAGGCCACGACGACCAGCGCCGAGGCGGACACAACCCGACGCCG 670

Db 42 AAACGCCCGGACCCGATACACACGACCGCGCTCCCGGCTCCCGACCCGA 101

671 TCACTGCAACCAAGCGCTCCGAAC TTGAAGCGGAACACTTACCTTCCCTAGAA 730

Db 102 ACCCGGATCCGGCGACCCGAAACCGACACCTCTCGAACTCCGACTCCCCGGCGCAA 161

731 ATGGCGTGATCATTACGAACCGACACCCGCAACGAATAATCAACGTTACTGTCCGTC 790

Db 162 ATAACTAACCGACTACGAACCGCGCCGGAACGAAAAAACACCGTACGCTC 221

791 TCGGACAATGAGCCCTACGCTAA 814

Db 222 GCCCGCAAAACGGCCTCCGACAA 245

[illegible]

Search completed: July 1, 2003, 14:40:22  
Job time : 343 secs

Job time : 343 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 14:16:19 ; Search time 1895 Seconds

(without alignments)  
9307.063 Million cell updates/sec

Title: US-08-468-190C-1\_COPY\_9874\_10962

Sequence: 1 ATGGCATCGCTACTGTGAAC.....CCGACTCATATATAATGCA 1089

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estic:\*  
9: gp\_estcl:\*  
10: gp\_estcl2:\*  
11: gp\_estcl3:\*  
12: gp\_estcl4:\*  
13: gp\_estcl5:\*  
14: gp\_estcl6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gp\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vtc:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1        | 43.2  | 4.9         | 493    | 12    | BE777414    |
| C 2        | 43.2  | 3.9         | 681    | 9     | AU076048    |
| C 3        | 41.4  | 3.8         | 591    | 9     | AU068783    |
| C 4        | 41.4  | 3.8         | 765    | 17    | AZ132997    |
| C 5        | 41.4  | 3.8         | 1053   | 17    | CNS060QTI   |
| C 6        | 38.2  | 3.5         | 575    | 13    | B1352860    |

|      |      |     |      |    |           |
|------|------|-----|------|----|-----------|
| C 7  | 37.8 | 3.5 | 925  | 17 | CNS0091P  |
| C 8  | 37   | 3.4 | 550  | 13 | BM596053  |
| C 9  | 37   | 3.4 | 675  | 13 | BM603983  |
| C 10 | 37   | 3.4 | 731  | 13 | BM585009  |
| C 11 | 36.6 | 3.4 | 409  | 9  | AT455021  |
| C 12 | 36.6 | 3.4 | 487  | 9  | AA803404  |
| C 13 | 36.6 | 3.4 | 487  | 9  | AA941430  |
| C 14 | 36.6 | 3.4 | 495  | 13 | AT395118  |
| C 15 | 36.6 | 3.4 | 538  | 13 | BT356698  |
| C 16 | 36.6 | 3.4 | 558  | 13 | BT242364  |
| C 17 | 36.6 | 3.4 | 559  | 9  | AA941412  |
| C 18 | 36.6 | 3.4 | 573  | 13 | BT166941  |
| C 19 | 36.6 | 3.4 | 588  | 9  | AA540589  |
| C 20 | 36.6 | 3.4 | 611  | 13 | BT484946  |
| C 21 | 36.6 | 3.4 | 616  | 13 | BT353787  |
| C 22 | 36.6 | 3.4 | 621  | 13 | BT169795  |
| C 23 | 36.6 | 3.4 | 625  | 9  | AT151547  |
| C 24 | 36.6 | 3.4 | 632  | 13 | AA948960  |
| C 25 | 36.6 | 3.4 | 632  | 13 | BT580110  |
| C 26 | 36.6 | 3.4 | 648  | 9  | AT151258  |
| C 27 | 36.6 | 3.4 | 652  | 9  | AA820152  |
| C 28 | 36.6 | 3.4 | 738  | 9  | AA391792  |
| C 29 | 36.6 | 3.4 | 742  | 9  | AT259265  |
| C 30 | 36.6 | 3.4 | 778  | 9  | AA539582  |
| C 31 | 36.6 | 3.3 | 1580 | 13 | BT219863  |
| C 32 | 36.2 | 3.3 | 1101 | 17 | CNS0000D1 |
| C 33 | 36   | 3.3 | 522  | 13 | BM317560  |
| C 34 | 36   | 3.3 | 538  | 12 | BT292594  |
| C 35 | 36   | 3.3 | 617  | 9  | AT657056  |
| C 36 | 36   | 3.3 | 624  | 14 | BO294678  |
| C 37 | 36   | 3.3 | 980  | 17 | CNS01250  |
| C 38 | 35.8 | 3.3 | 384  | 17 | BT7530    |
| C 39 | 35.6 | 3.3 | 379  | 17 | AT218897  |
| C 40 | 35.6 | 3.3 | 421  | 9  | AA695350  |
| C 41 | 35.6 | 3.3 | 421  | 9  | BT231965  |
| C 42 | 35.6 | 3.3 | 709  | 14 | BO840906  |
| C 43 | 35.4 | 3.3 | 367  | 13 | BM448221  |
| C 44 | 35.4 | 3.3 | 482  | 12 | BT040664  |
| C 45 | 35.4 | 3.3 | 489  | 13 | BM447647  |

#### ALIGNMENTS

RESULT 1  
LOCUS BE777414/C 493 bp mRNA linear EST 20-SEP-2000  
DEFINITION MY-35-A-04 Pinfestansmy Phytophthora infestans cDNA, mRNA sequence.  
ACCESSION BE777414  
VERSION BE777414.1 GI:10231053  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Kamoun, S., Hrabner, P. T., Sobral, R. W. S., Nus, D. and Govers, F.  
TITLE Initial assessment of gene diversity for the complete pathogen  
JOURNAL Phytophthora infestans based on expressed sequences  
MEDLINE Fungal Genet. Biol. 28 (2), 94-106 (1999)  
COMMENT  
Contact: Govers F  
Laboratory of Phytopathology  
Wageningen University  
Bilthoven 9, P.O. Box 8025, 6700 EE, Wageningen, The Netherlands  
Tel: 31 317 483 138  
Fax: 31 317 483 412  
Email: francine.govers@medew.fyto.wau.nl.  
Location/Qualifiers  
1. 493  
/organism="Phytophthora infestans"  
/strain="DDR7602, AI mating type"  
/db\_xref="taxon:4787"

BASE COUNT  
ORIGIN

87 a 128 c 183 g 95 t

|                          |       |                |          |            |
|--------------------------|-------|----------------|----------|------------|
| Query Match              | 4.08  | Score 43.2     | DB 12    | Length 493 |
| Best Local Similarity    | 56.28 | Pred. No. 0.23 |          |            |
| Matches 81, Conservative | 0     | Mismatches 63  | Indels 0 | Gaps 0     |

548 TCGGATTACCTACCGTCCACTTGAGGCTATCTCAGGACCGAGGAAGTTGGCGCACT 607

608 GGCAAGCGTACGTTGCCACGAGGCCACGACGACCGCCGAGGGGACACCCCGGACGC 667

Db 99 GGCTGGGGAGCCGCCACGACCACGGCCACCTCGGCCACCTCCACGTCCA 40

Db 39 CCGCGTCCTCCACAGGAGCGCTCCG 16

RESULT 2  
AU076048/C

DEFINITION  
AB078048 Rice panicle at flowering stage Oryza sativa (japonica  
cultivar-group) cDNA clone E2735\_6Z, mRNA sequence.

| KEYWORDS | EST.                                     |
|----------|--|
| SOURCE   | Oruva sativa (Japansica cultivar-around) |

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

**AUTHORS**  
Sasaki, T. and Yamamoto, K.  
Rice, CDNA from particles at flowering stage

COMMENT  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources

202 0002, Japan  
Tel: 01-298-38-7441  
Fax: 01-298-38-7460

PROJECT = 'RGP'.

Location/Qualifiers

```

/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"

```

```
/clone_1lb="Rice panicle at flowering stage"  
/dev stnne="flowering stage"
```

|            | stage <sup>a</sup> |       |       |       |          |
|------------|--------------------|-------|-------|-------|----------|
| BASE COUNT | 141 a              | 171 c | 191 g | 174 t | 4 others |

Query Match 3.98; Score 42; DB 9; Length 681;

560 CCGTCCAACTTTGAGGCTTATCTTCAGAGCCGAGGAAAGTTTGGGCGCAACTTGGCCAAAGCCGACG 619

Db 354 CCATCCGGGTTTCAGAGCTGCGCTTTGGAGATGTGTACGTGGCGCAGCGCGCTGGCCATGT 295

OY 620 TTGCCAGGAGGCCACGAGACCAAGCGCCGAGCGCAACCCCGAGCGCCGTACTGCAA 679

Db 294 CGGCCACCGAGCCCGCCCGGAGACAGGACGACGACGCGCGCGGCGAGCTTGAGGCGCT 235

OY 680 CCAGGCGCTCGGAATTGAGCGGAACATTATTCCTTCCTG 722

Db 234 GCAGGGCCACCCACTTCCGGGCTGCGCTCTGTGAACTTGCCTG 192

| RESULT 3   |  |
|------------|--|
| AU068783/c |  |
| LOCUS      | 591 bp   |
| DEFINITION | mRNA linear EST 01-APR-2002                                      |
|            | AU068783 Rice callus Oryza sativa (japonica cultivar-group) cDNA |

|           |            |
|-----------|------------|
| ACCESSION | AU068783   |
| VERSION   | AU068783.1 |
|           | GI:5003634 |

| REFERENCE | SOURCE                                  | ORGANISM                               |
|-----------|---|--|
| 1         | Oryza sativa (japonica cultivar-group). | Oryza sativa (japonica cultivar-group) |

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

**AUTHORS** Idimamullo, A. and Sasaki, I.  
**TITLE** Rice cDNA from callus (1998)  
**ABSTRACT** *Abstract text is missing from the page.*

National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

**Tel: 81-298-38-7441**  
**Fax: 81-298-38-7468**

| FEATURES | Location/Qualifiers. |
|----------|----------------------|
| 1        | 501                  |

```
/cultivar="Nipponbare"  
/db_xref="taxon:39947"
```

```

/crome_id= nice carus
/note="Vector: plusescript II SK+; Site_1: SalI; Site_2:
      not inserted from the cell line by using

```

| BASE COUNT | of pBluescript II SK+ phagemid. | "        |
|------------|---------------------------------|----------|
| 137 a      | 139 c                           | 159 g    |
|            |                                 | 155 t    |
|            |                                 | 1 others |

|             |       |             |        |             |
|-------------|-------|-------------|--------|-------------|
| Query Match | 3.88; | Score 41.4; | DB: 9; | Length 591; |
|-------------|-------|-------------|--------|-------------|

560 CCGTCCACCTTGAGGCTATCTCAGGACCGAGGAAGTTGGCGCACTGGCAAGCGTACG 619

DB 260 CCATCCGGGTTGAGAGCTGGCTTTGGAGAGATGATGACGTGGGCGAGGGCGTGGCCCATGT 201

Db 200 CGGCCACGAGCCACCGCCGGGACAGGCACACGAGCGCGCGACGTTGACGCCCT 141

Db 140 GCAAGCCACCCACTTCCGGCTGCTCTGCACCTTGGCTG 98

## RESULT 4

|            |  |
|------------|--|
| DEFINITION | OSJNBb0108G22f CUGI Rice BAC Library (ECORI) <i>Oryza sativa</i> genomic clone OSJNBb0108G22f. DNA sequence. |
|------------|--|









Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)"

BASE COUNT 182 a 173 c 244 g 132 t

Query Match 3.4%; Score 37; DB 13; Length 731;

Best Local Similarity 53.0%; Pred. No. 17;

Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

543 CACCTTCGATTACCTACCGCTTCCATGAGCTATCTCAGAGACCGAGAAAGTTGGCG 602

459 CACCTTCGCTTTGTTTATACCACTTTTGGCCGCCACTGTACCGCGGTTTGGT 400

603 CAACTGGCAAGGATGATGTCACGAGGACACGAGCAGCGCCGCGGACACACACCC 662

399 TACCACTCTCTGTCGTAGCCGCCGCCACACACGACGCGCGGTTGTCTACAC 340

663 GACGCGCCCTACTGACACGCGCCCTCCG 691

339 CGTACCCGCCACCGTCACACGCGCTCCG 311

RESULT 11

LOCUS A1455021/c 409 bp mRNA linear EST 19-APR-2001

DEFINITION LD04335.5prime LD Drosophila melanogaster embryo Bluescript

Drosophila melanogaster cDNA clone LD04335 5prime, mRNA sequence.

ACCESSION A1455021

VERSION A1455021.1 GI:4298123

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 409)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HHMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

Plate: 43 row: C column: 11

High quality sequence stop: 363.

Location/Qualifiers

1..409

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="LD04335"

/clone\_1lb="LD Drosophila melanogaster embryo Bluescript"

/sex="male and female"

/dev\_stage="0 to 24 hours mixed stage embryonic"

/lab\_host="SOLR"

/note="Organ: embryo; Vector: Bluescript SK; Site: 1; EcORI

Synthesis kit. Oligo dt-primed and directionally cloned at

EcORI and XhoI in Bluescript SK(+/-)"

BASE COUNT 127 a 88 c 128 g 66 t

ORIGIN

Query Match 3.4%; Score 36.6; DB 9; Length 409;

Best Local Similarity 62.6%; Pred. No. 16;

Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

29 TCCTTGGCGGACGCTTCGACCCCTTCGCGGATGGAGTCGTACTGGAATACG 88

281 TCGTTTTCGCGCTTCTTGGCGGATGGAGTCGTACTGGAATACGTCGTC 222

89 TCTCCGACGATGACGACGATCAGTCATCTG 119

Db 221 TCCTGATGGAGTTCGTGATCAGATCTT 191

RESULT 12

AA803404 487 bp mRNA linear EST 23-APR-2001

DEFINITION GM10781.5prime GM Drosophila melanogaster ovary Bluescript

Drosophila melanogaster cDNA clone GM10781 5prime, mRNA sequence.

ACCESSION AA803404

VERSION AA803404.1 GI:2872523

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 487)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HHMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

Plate: 107 row: G column: 9

High quality sequence stop: 380.

Location/Qualifiers

1..487

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GM10781"

/clone\_1lb="GM Drosophila melanogaster ovary Bluescript"

/sex="female"

/dev\_stage="newly eclosed females; germlinum-stage 6"

/lab\_host="SOLR"

/note="Organ: ovary; Vector: Bluescript SK; Site: 1; EcORI;

Site: 2; XhoI: Constructed using Stratagene ZAP-cDNA

Synthesis kit. Oligo dt-primed and directionally cloned at

EcORI and XhoI in Bluescript SK(+/-)"

BASE COUNT 143 a 104 c 151 g 88 t 1 others

ORIGIN

Query Match 3.4%; Score 36.6; DB 9; Length 487;

Best Local Similarity 62.6%; Pred. No. 18;

Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

29 TCCTTGGCGGACGCTTCGACCCCTTCGCGGATGGAGTCGTACTGGAATACG 88

312 TCGTTTTCGCGCTTCTTGGCGGATGGAGTCGTACTGGAATACGTCGTC 253

89 TCTCCGACGATGACGACGATCAGTCATCTG 119

252 TCCTGATGGAGTTCGTGATCAGATCTT 222

RESULT 13

AA941430 487 bp mRNA linear EST 19-APR-2001

LOCUS LD25565.5prime LD Drosophila melanogaster embryo POT2 Drosophila

melanogaster cDNA clone LD25565 5prime, mRNA sequence.

ACCESSION AA941430

VERSION AA941430.1 GI:3101343

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 487)

AUTHORS Harvey, D. J., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S. and Rubin, G. M.  
TITLE BDGP/HMT Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT Contact: Stapleton, M.

BDGP Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: 255 row: F column: 5  
High quality sequence stop: 436.  
Location/Qualifiers

## FEATURES

source

1. 487  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="LD25565"  
/clone\_1lb="LD Drosophila melanogaster embryo pot2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XDL Blue"  
/note="Organ: embryo; Vector: pOT2; Site: 1; EcoRI; Site: 2;  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2."  
BASE COUNT 143 a 107 c 150 g 87 t  
ORIGIN

Query Match 3.4%; Score 36.6; DB 9; Length 487;  
Best Local Similarity 62.6%; Pred. No. 18;  
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 29 TCCTTCCCGCAGCCTCCGACCCCTTCGCGCGGAGTGCATGTCGATCGAATCAACG 88  
Dp 320 TCGTTTTCGCGCTTCCTTCGCGCGGAGTGCATGTCGATCGAATCAACG 261  
OY 89 TCTCCGCGAGATTGACGACGATCAATCGT 119  
Db 260 TCCCTGATGAGATTGCTGCGATCGATCTT 230

RESULT 14  
AI295118/c 495 bp mRNA linear EST 19-APR-2001  
LOCUS LP08655.Sprime LP Drosophila melanogaster larval-early pupal pOT2  
DEFINITION Drosophila melanogaster cDNA clone LP08655 Sprime, mRNA sequence.  
ACCESSION AI295118  
VERSION AI295118.1 GI:3944525  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 495)  
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S. and Rubin, G. M.  
BDGP/HMT Drosophila EST Project  
Unpublished (2001)  
Contact: Stapleton, M.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BDGP Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: 86 row: E column: 7  
High quality sequence stop: 493.  
Location/Qualifiers

## FEATURES

source

1. 495  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="LP08655"  
/clone\_1lb="LP Drosophila melanogaster larval-early pupal  
pOT2."

/sex="male and female"  
/dev\_stage="larvae, pupae"  
/lab\_host="DH5-alpha"  
/note="Organ: whole body; Vector: pOT2; Site: 1; EcoRI;  
Site: 2; XhoI; Sized fractionated cDNAs were directly  
ligated into pOT2. Plasmid cDNA library."

BASE COUNT 147 a 108 c 153 g 87 t  
ORIGIN

Query Match 3.4%; Score 36.6; DB 9; Length 495;  
Best Local Similarity 62.6%; Pred. No. 18;  
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 29 TCCTTCCCGCAGCCTCCGACCCCTTCGCGCGGAGTGCATGTCGATCGAATCAACG 88  
Dp 323 TCGTTTTCGCGCTTCCTTCGCGCGGAGTGCATGTCGATCGAATCAACG 264  
OY 89 TCTCCGCGAGATTGACGACGATCAATCGT 119  
Db 263 TCCCTGATGAGATTGCTGCGATCGATCTT 233

RESULT 15  
B1356698/c 538 bp mRNA linear EST 31-JUL-2001  
LOCUS R42893.Sprime RE Drosophila melanogaster normalized Embryo pFIC-1  
DEFINITION Drosophila melanogaster cDNA clone R42893 5 similar to spIC6;  
Fban001225 structural protein located on: X 5E1-5E1:: 05/13/2001  
RNA sequence.  
ACCESSION B1356698  
VERSION B1356698.1 GI:15051152  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 538)  
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson  
J., Chamez, M., Chavez, C., Dorsett, V., Partan, D., Frise, E., George  
R., Gonzalez, M., Guatin, H., Harris, N., Li, P., Liao, G., Mista, S.,  
Mingall, C. J., Nuno, J., Pacleb, J., Paragas, V., Park, S.,  
Phoumenavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin  
G. M.  
BDGP/HMT RE Drosophila EST Project  
Unpublished (2001)  
Contact: Stapleton, M.

TITLE  
JOURNAL  
COMMENT

BDGP Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
hit genomic AC009437; arm X 1602150.6347167)  
estimated cyto:506-681; 05/13/2001  
Plate: RE:428 row: H column: 9  
High quality sequence stop: 459.  
Location/Qualifiers

## FEATURES

source

1. 538  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="R42893"  
/clone\_1lb="RE Drosophila melanogaster normalized Embryo  
pFIC-1"  
/sex="male and female"  
/dev\_stage="0-24 hours mixed stage embryonic"  
/lab\_host="DH5-alpha Tona"  
/note="Organ: embryo; Vector: pFIC1; Site: 1; XhoI; Site: 2;  
BamHI; Library was kindly generated by Piero Carninci at  
the RIKEN. The library was normalized and excised using  
Cre recombinase. Plasmid cDNA library."

BASE COUNT 154 a 117 c 170 g 97 t  
ORIGIN

|                          |        |                |           |             |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match              | 3.48;  | Score 36.6;    | DB 13;    | Length 538; |
| Best Local Similarity    | 62.68; | Pred. NO. 19;  |           |             |
| Matches 57; Conservative | 0;     | Mismatches 34; | Indels 0; | Gaps 0;     |

|    |     |  |     |
|----|-----|--|-----|
| Oy | 29  | TCCTTCCGCGACGCTCGACACCTTTCGGGCGCATGGATCGATCGAATACAG        | 88  |
|    |     |  |     |
| Db | 312 | TGCTTTTGGCGCTTCTTGGCGCTGGCGAGCCATCGATCTGTAAACACTGCCATGCTCC | 253 |
| Oy | 89  | TCTCCGACGATTCGACGATTCATACGT                                | 119 |
|    |     |  |     |
| Db | 252 | TCCTCGATGGGATTTCTGTCGATTCAGATCTT                           | 222 |

Search completed: July 1, 2003, 16:03:52  
Job time : 1903 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 14:30:58; Search time 70 Seconds  
(without alignments) 4771.014 Million cell updates/sec

Title: US-08-468-190C-1\_COPY\_9874\_10962

Perfect score: 1089

Sequence: 1 ATGCATCGCTACTTGAGAC.....CCGACTCAATATAAATGTGA 1089

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Optimal number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents -NA:\*  
1: /cgn2\_6/ptodata/1/ins/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ins/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ins/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ins/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ins/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ins/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length  | DB ID | Description         |
|------------|--------|-------------|---------|-------|---------------------|
| 1          | 1089   | 100.0       | 13473   | 5     | PCT-US96-03916-1    |
| 2          | 1087.4 | 99.9        | 18912   | 5     | PCT-US96-03916-59   |
| 3          | 346    | 31.8        | 534     | 5     | PCT-US96-03916-18   |
| 4          | 38.2   | 3.5         | 4403765 | 4     | US-09-103-840A-2    |
| 5          | 38.2   | 3.5         | 4411529 | 4     | US-09-103-840A-1    |
| 6          | 35.2   | 3.2         | 19718   | 4     | US-08-961-527-99    |
| 7          | 34.6   | 3.2         | 50341   | 1     | US-08-247-901C-1    |
| 8          | 34.6   | 3.2         | 50341   | 2     | US-09-075-904-1     |
| 9          | 34.6   | 3.2         | 52297   | 4     | US-09-426-436-1     |
| 10         | 34.6   | 3.0         | 52297   | 4     | US-08-705-557-1     |
| 11         | 32.2   | 3.0         | 4403765 | 4     | US-09-103-840A-2    |
| 12         | 32.2   | 3.0         | 4411529 | 4     | US-09-103-840A-1    |
| 13         | 31.8   | 2.9         | 1145    | 3     | US-08-809-103B-7    |
| 14         | 31.8   | 2.9         | 1148    | 3     | US-08-809-103B-1    |
| 15         | 31.8   | 2.9         | 1150    | 3     | US-08-809-103B-3    |
| 16         | 31.8   | 2.9         | 1150    | 3     | US-08-809-103B-5    |
| 17         | 31.6   | 2.9         | 7218    | 1     | US-08-232-63-14     |
| 18         | 31.6   | 2.9         | 843     | 3     | US-08-513-974B-375  |
| 19         | 31.6   | 2.9         | 25165   | 4     | US-09-453-702B-39   |
| 20         | 30.8   | 2.8         | 948     | 4     | US-09-134-001C-1118 |
| 21         | 30.8   | 2.8         | 1728    | 1     | US-08-477-097-1     |
| 22         | 30.8   | 2.8         | 1728    | 2     | US-08-878-957-1     |
| 23         | 30.8   | 2.8         | 1752    | 1     | US-08-427-097-15    |
| 24         | 30.8   | 2.8         | 1752    | 2     | US-08-878-957-15    |
| 25         | 30.8   | 2.8         | 1752    | 2     | US-08-878-957-15    |
| 26         | 30.8   | 2.8         | 1752    | 2     | US-08-878-957-15    |
| 27         | 30.4   | 2.8         | 3467    | 1     | US-07-745-206A-12   |

|    |      |     |      |   |                   |                   |
|----|------|-----|------|---|-------------------|-------------------|
| 28 | 30.4 | 2.8 | 5467 | 2 | US-08-311-363-12  | Sequence 12, Appl |
| 29 | 30   | 2.8 | 843  | 4 | US-08-339-214-7   | Sequence 7, Appl  |
| 30 | 30   | 2.8 | 843  | 4 | US-09-006-755B-2  | Sequence 2, Appl  |
| 31 | 30   | 2.8 | 846  | 3 | US-08-815-190A-1  | Sequence 1, Appl  |
| 32 | 30   | 2.8 | 924  | 4 | US-08-339-214-33  | Sequence 33, Appl |
| 33 | 30   | 2.8 | 924  | 4 | US-08-339-214-33  | Sequence 34, Appl |
| 34 | 30   | 2.8 | 924  | 4 | US-08-339-214-35  | Sequence 35, Appl |
| 35 | 30   | 2.8 | 972  | 4 | US-09-479-524-1   | Sequence 1, Appl  |
| 36 | 30   | 2.8 | 972  | 4 | US-09-479-524-7   | Sequence 7, Appl  |
| 37 | 30   | 2.8 | 1790 | 2 | US-08-810-453-1   | Sequence 1, Appl  |
| 38 | 30   | 2.8 | 1841 | 5 | PCT-US95-00362-1  | Sequence 1, Appl  |
| 39 | 30   | 2.8 | 1890 | 4 | US-09-290-640-24  | Sequence 24, Appl |
| 40 | 30   | 2.8 | 2432 | 4 | US-08-339-214-29  | Sequence 29, Appl |
| 41 | 30   | 2.8 | 3423 | 4 | US-08-471-112A-2  | Sequence 11, Appl |
| 42 | 30   | 2.8 | 5430 | 3 | US-09-012-515A-11 | Sequence 11, Appl |
| 43 | 30   | 2.8 | 5430 | 3 | US-08-360-144A-11 | Sequence 11, Appl |
| 44 | 30   | 2.8 | 5430 | 3 | US-09-012-504A-11 | Sequence 11, Appl |
| 45 | 30   | 2.8 | 7653 | 4 | US-08-471-112A-1  | Sequence 1, Appl  |

## ALIGNMENTS

RESULT 1  
PCT-US96-03916-1  
Sequence 1, Application PC/TUS9603916  
GENERAL INFORMATION:  
APPLICANT: Wild, Martha A.  
APPLICANT: Cochran, Mark D.  
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESS: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/03916  
FILING DATE: 23-MAR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,597  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 39116-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 391-0525  
TELEFAX: (212) 278-0400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13473 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1059..2489  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2575..4107  
FEATURE:

```

NAME/KEY: CDS
LOCATION: 4113..4445
FEATURE:
NAME/KEY: CDS
LOCATION: 4609..5487
FEATURE:
NAME/KEY: CDS
LOCATION: 5697..8654
FEATURE:
NAME/KEY: CDS
LOCATION: 9874..10962
FEATURE:
NAME/KEY: CDS
LOCATION: 11159..12658
FEATURE:
NAME/KEY: CDS
LOCATION: 12665..13447
PCT-US96-03916-1

```

```

Query Match 100.0%; Score 1089; DB 5; Length 13473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1089; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGGCATCGCTACTTGGAACTCTGGCTCTCTCTCCCGGAGAGCTCGACCCCTTGGCGCG 60
DB 9874 ATGGCATCGCTACTTGGAACTCTGGCTCTCTCTCCCGGAGAGCTCGACCCCTTGGCGCG 9933
QY 61 ATGGGAATCGTATCTAGTGAATCAGTCTCCGCGAGAGATGACAGATCAGATCGTG 120
DB 9934 ATGGGAATCGTATCTAGTGAATCAGTCTCCGCGAGAGATGACAGATCAGATCGTG 9993
QY 121 ATGCTCGGCTCGCCCGGAGAGTACATTAATTCAGTATTTTCATGCTGGCCAG 180
DB 9994 ATGCTCGGCTCGCCCGGAGAGTACATTAATTCAGTATTTTCATGCTGGCCAG 10053
QY 181 AGACCCGAGAAACCTACTAGAGAACGCTCGCGCTTGGCTGATATTAACAAC 240
DB 10054 AGACCCGAGAAACCTACTAGAGAACGCTCGCGCTTGGCTGATATTAACAAC 10113
QY 241 CAGTGTACAGAGAACTTAGCGAGAGAGCGCTTTGAAAATTGCACTCATCGCTTCT 300
DB 10114 CAGTGTACAGAGAACTTAGCGAGAGAGCGCTTTGAAAATTGCACTCATCGCTTCT 10173
QY 301 GTTTTGTGGCTGTAAAGTACGAGTACAGCTTCTCCGCTTGAACAGACTAACCGGA 360
DB 10174 GTTTTGTGGCTGTAAAGTACGAGTACAGCTTCTCCGCTTGAACAGACTAACCGGA 10233
QY 361 CCTCCAGACCCGTTAAGCTACATACGAATCTCTGCTCCGAGAGAGAGGAGATGTC 420
DB 10234 CCTCCAGACCCGTTAAGCTACATACGAATCTCTGCTCCGAGAGAGAGGAGATGTC 10293
QY 421 TACGTAATGTGGCTAGAGAGACAGCAACCAATGAGCTTCGCGATCAACTA 480
DB 10294 TACGTAATGTGGCTAGAGAGAGACAGCAACCAATGAGCTTCGCGATCAACTA 10353
QY 481 TCGGTATCAATTCGCGAACACCGCGGAGCTCGCGAGCTATATTCGAAGGCTTGT 540
DB 10354 TCGGTATCAATTCGCGAACACCGCGGAGCTCGCGAGCTATATTCGAAGGCTTGT 10413
QY 541 CGGACCTTGGATTAATCACTTACCGTCACTTGAAGGCTATCTCAGAGAGAGAAAGTTGG 600
DB 10414 CGGACCTTGGATTAATCACTTACCGTCACTTGAAGGCTATCTCAGAGAGAGAAAGTTGG 10473
QY 601 CGCAACTGGAAGAGTGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 10474 CGCAACTGGAAGAGTGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10533
QY 661 CGGAGCGCGCTCACTGACAGAGCGCTCGGAATTTGAAGCGGAACCTTACCTTCC 720
DB 10534 CGGAGCGCGCTCACTGACAGAGCGCTCGGAATTTGAAGCGGAACCTTACCTTCC 10593
QY 721 TGGGTAGAAAATGGCGTGGATCATTAAGAACGAGACCCGGAACGAATTTCAAGCTT 780
DB 10594 TGGGTAGAAAATGGCGTGGATCATTAAGAACGAGACCCGGAACGAATTTCAAGCTT 10653

```

```

DB 10594 TGGGTAGAAAATGGCGTGGATCATTAAGAACGAGACCCGGAACGAATTTCAAGCTT 10653
QY 781 ACTGTCCGCTCGGAGAACATGAGCCCTACGCTAATTTGGGGTAAACGTTGGCTGCCGTGTG 840
DB 10654 ACTGTCCGCTCGGAGAACATGAGCCCTACGCTAATTTGGGGTAAACGTTGGCTGCCGTGTG 10713
QY 841 AGCGCAAGATCGGCTCGTCTGATTTGATTTTCATCTGCAACCAAAATGTCGACCCCG 900
DB 10714 AGCGCAAGATCGGCTCGTCTGATTTGATTTTCATCTGCAACCAAAATGTCGACCCCG 10773
QY 901 CACCGAAATTTAGACAGAGTCTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 10774 CACCGAAATTTAGACAGAGTCTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10833
QY 961 TCGGGAATTTGAGACCATGCTTGGCGGAATTAACAAGGCGGTGACAGATAGT 1020
DB 10834 TCGGGAATTTGAGACCATGCTTGGCGGAATTAACAAGGCGGTGACAGATAGT 10893
QY 1021 GAACCTTGGAACTGTTGGATTTGTTAACCCGCTCTGCGCTAAGCTGCGCCGACTCAATA 1080
DB 10894 GAACCTTGGAACTGTTGGATTTGTTAACCCGCTCTGCGCTAAGCTGCGCCGACTCAATA 10953
QY 1081 AAAATGGA 1089
DB 10954 AAAATGGA 10962

RESULT 2
PCT-US96-03916-59
Sequence 59, Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,597
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39116-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 18912 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS

```





STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/03916  
FILING DATE: 23-MAR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,597  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 39116-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 534 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..533  
PCT-US96-03916-18

Query Match 31.8%; Score 346; DB 5; Length 534;  
Best Local Similarity 100.0%; Pred. No. 7.3e-98;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 744 TTACGACCGACCGACCGCAACGAAATTCACAGTTCCTCGCTCGGACAAATGAG 803  
|||  
Db 534 TTACGACCGACCGACCGCAACGAAATTCACAGTTCCTCGCTCGGACAAATGAG 475  
|||  
QY 804 CCCACGCTAATTTGGGTTAAACCGTGGCTGCGTGTGAGGCAACGATCGGCTGTCAT 863  
|||  
Db 474 CCCACGCTAATTTGGGTTAAACCGTGGCTGCGTGTGAGGCAACGATCGGCTGTCAT 415  
|||  
Db 864 TGTATTTTCATCTCTACACGAAACATGTGACACCCCGACCGAATTTAGACAGGTCCTC 923  
|||  
Db 414 TGTATTTTCATCTCTACACGAAACATGTGACACCCCGACCGAATTTAGACAGGTCCTC 355  
|||  
QY 924 GCAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 983  
|||  
Db 354 GCAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 295  
|||  
QY 984 TCGCTGGCAATTAACAAGGCGCTGACACGAGATAGTGAACCTTGTGAACTGTTGGCAT 1043  
|||  
Db 294 TCGCTGGCAATTAACAAGGCGCTGACACGAGATAGTGAACCTTGTGAACTGTTGGCAT 235  
|||  
QY 1044 TGTTAACCGCTCTGCTAGAGCTGCGCCGACCTCAATTAATGATGCA 1089  
|||  
Db 234 TGTTAACCGCTCTGCTAGAGCTGCGCCGACCTCAATTAATGATGCA 189  
|||

RESULT 4  
US-09-103-840A-2/c  
Sequence 2, Application US/09103840A

GENERAL INFORMATION:  
PATENT NO. 6294328  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2

Query Match 3.5%; Score 38.2; DB 4; Length 4403765;  
Best Local Similarity 58.3%; Pred. No. 1.5;  
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 29 TCCCTGGCGGACGCTGCGACACCTTGGCGGATGGGAATCGATCATGGAATACAG 88  
|||  
Db 619394 TCCCGCGCGACGCTGCGACACCTTGGCGGATGGGAATCGATCATGGAATACAG 619335  
|||  
QY 89 TCTCCGCGACGATGACGACGATCATGCTGATGCTGCGGCTCGCCCGGAAGC 143  
|||  
Db 619334 TCGGCCCGACGACGATGCTGCTGCGGCGGAGCGGGAACACCGCGCAAGACCGCGCC 619280  
|||

RESULT 5  
US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
PATENT NO. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 3.5%; Score 38.2; DB 4; Length 4411529;  
Best Local Similarity 58.3%; Pred. No. 1.5;  
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 29 TCCCTGGCGGACGCTGCGACACCTTGGCGGATGGGAATCGATCATGGAATACAG 88  
|||  
Db 618034 TCCCGCGCGACGCTGCGACACCTTGGCGGATGGGAATCGATCATGGAATACAG 617975  
|||  
QY 89 TCTCCGCGACGATGACGACGATCATGCTGATGCTGCGGCTCGCCCGGAAGC 143  
|||  
Db 617974 TCGGCCCGACGACGATGCTGCTGCGGCGGAGCGGGAACACCGCGCAAGACCGCGCC 617920  
|||

RESULT 6  
US-08-961-527-99  
Sequence 99, Application US/08961527

GENERAL INFORMATION:  
PATENT NO. 6420135  
APPLICANT: CHARLES KUNSCH



OY 611 AACG 615  
18485 AACG 18481

## RESULT 8

US-09-075-904-1/c  
Sequence 1, Application US/09075904  
Patent No. 5994137  
GENERAL INFORMATION:  
APPLICANT: Jacobs, et al.  
TITLE OF INVENTION: L5 SHUTTLE PHASMIDS  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amster, Rothstein & Epstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (ASCII)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/075.904  
FILING DATE: May 11, 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/247.901  
FILING DATE: May 23, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bogosian, Elizabeth A  
REGISTRATION NUMBER: 39,911  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50341  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: L5 shuttle phasmid sequence  
HYPOTHETICAL: NO  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: L5 mycobacteriophage  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION: No. 5994137e  
AUTHORS:  
TITLE:  
JOURNAL:

VOLUME:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-09-075-904-1

Query Match 3.2% Score 34.6; DB 2; Length 50341;  
Best Local Similarity 49.2%; Pred. No. 2.9;  
Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

OY 431 TTCGGCTAGACGACCAAGACCAATTGACGCTTCGCGATCCAACTATCGGTATC 490  
18665 TTCCGCTGTGAGATGATGACACCCCAAGACGATGTCTGCTCAGGGCTCGGTGGAC 18606  
OY 491 AATTCCGACACCGCCGCGACCTGCGGACTCTATTCACAGGCTTCGTCACCTTCG 550  
18605 AAGTAGCGAGCGCCGCGGTAGTTCGCCGCTCTTCAACACGACCTTGACGGAGGTCG 18546  
OY 551 GATTACCTACCGTCCCACTTATGAGGCTTATCTCAGACCGAGGAAGTTGGCGCACTGGC 610  
18545 TAGAACAACCCCTCCAGCTGTGTGCGCAGTACACACCCCTGTCACCGGTCTCAAGTTG 18486  
OY 611 AACG 615  
18485 AACG 18481

## RESULT 9

US-09-426-436-1/c  
Sequence 1, Application US/09426436  
Patent No. 6225066  
GENERAL INFORMATION:  
APPLICANT: William R. Jacobs, Jr.  
APPLICANT: Barry R. Bloom  
ADDRESSEE: Graham F. Hatfull  
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC  
TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amster, Rothstein & Epstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (ASCII)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/426.436  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/705.557  
FILING DATE:  
APPLICATION NUMBER: US/08/057.531  
FILING DATE:  
APPLICATION NUMBER: 07/833.431  
FILING DATE: February 7, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pasqualini, Patricia A.  
REGISTRATION NUMBER: 34,894  
REFERENCE/DOCKET NUMBER: 96700/238  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:

```

1 SEQUENCE CHARACTERISTICS:
2 LENGTH: 52297
3 TYPE: nucleotide
4 STRANDEDNESS: single
5 TOPOLOGY: linear
6 MOLECULE TYPE:
7 DESCRIPTION: phage genome sequence
8 HYPOTHEICAL: no
9 ANTI-SENSE: no
10 FRAGMENT TYPE: not applicable.
11 ORIGINAL SOURCE:
12 ORGANISM: mycobacteriophage L5
13 STRAIN: not applicable
14 INDIVIDUAL ISOLATE: L5
15 DEVELOPMENTAL STAGE: not applicable
16 HAPLOTYPE: not applicable
17 TISSUE TYPE: not applicable
18 CELL TYPE: not applicable
19 ORGANELLE: not applicable
20 IMMEDIATE SOURCE: mycobacteriophage L5 particles
21 POSITION IN GENOME: entire genome
22 FEATURE:
23 NAME/KEY:
24 LOCATION:
25 IDENTIFICATION METHOD:
26 OTHER INFORMATION:
27 PUBLICATION INFORMATION:
28 AUTHORS: Hatfull and Sarkis
29 TITLE: DNA Sequence, Structure and Gene
30 TITLE: Expression of Mycobacteriophage L5:
31 TITLE: A Phage System for Mycobacterial
32 TITLE: Genetics
33 JOURNAL: Molecular Microbiology
34 VOLUME: 7
35 PAGES: 395-405
36 DATES: 1993
37 US-09-426-436-1
38
39 Query Match 3 28: Score 34.6; DB 4; Length 52297;
40 Best Local Similarity 49.28; Pred. No. 2.9;
41 Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
42
43 Oy 431 TTCGGCTAGACGACACCAAGAACCATTGACGCTTCGCGATCCCACTATCGGTGATC 490
44 Db 18562 TTCGGCTAGACGACACCAAGAACCATTGACGCTTCGCGATCCCACTATCGGTGATC 185030
45 Oy 491 AATTCGCGAACCAGCCGCCGACTCGGAGCTCTATTCCAAAGGCTTCGTGTGCGACCTCG 550
46 Db 18502 AAGTAGCGAGCGCCGCCGGGTAGTTCGCCGCTCTTCAACAAGACCTTGACGGGAGGGTGC 184433
47 Oy 551 GATTACTACCGTCCCACTGAGGCGCTATCTCAGACGACGAGAAAGTTGGCGCACTGGC 610
48 Db 18442 TAGAAACAACCCCTCCAGCTGTGTGGCAGGTACACACACCTGGTCAACGCGTCGAAGTTG 183838
49 Oy 611 AAGCG 615
50 Db 18382 AAACG 18378
51
52 RESULT 10
53 US-08-705-557-1/c
54 Sequence 1, Application US/08705557
55 Patent No. 6300061
56 GENERAL INFORMATION:
57 APPLICANT: William R. Jacobs, Jr.
58 APPLICANT: Barry R. Bloom
59 APPLICANT: Graham F. Hatfull
60 TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
61 TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
62 NUMBER OF SEQUENCES: 1
63 CORRESPONDENCE ADDRESS:
64 ADDRESSEE: Amstet, Rotenstein & Edenstein

```

```

1 STREET: 90 Park Avenue
2 CITY: New York
3 STATE: New York
4 COUNTRY: U S.A.
5 ZIP: 10016
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
9 COMPUTER: IBM PC Compatible
10 OPERATING SYSTEM: MS-DOS
11 SOFTWARE: Word Processor (ASCII)
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/705,557
14 FILING DATE:
15 CLASSIFICATION:
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US/08/057,531
18 FILING DATE:
19 APPLICATION NUMBER: 07/833,431
20 FILING DATE: February 7, 1992
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Pasqualini, Patricia A.
23 REGISTRATION NUMBER: 34,894
24 REFERENCE/DOCKET NUMBER: 96700/238
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (212) 697-5995
27 TELEFAX: (212) 286-0854 or 286-0082
28 TELETYPE: TWX 710-581-4766
29 INFORMATION FOR SEQ ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 52297
32 TYPE: nucleotide
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE:
36 DESCRIPTION: phage genome sequence
37 HYPOTHEICAL: no
38 ANTI-SENSE: no
39 FRAGMENT TYPE: not applicable.
40 ORIGINAL SOURCE:
41 ORGANISM: mycobacteriophage L5
42 STRAIN: not applicable
43 INDIVIDUAL ISOLATE: L5
44 DEVELOPMENTAL STAGE: not applicable
45 HAPLOTYPE: not applicable
46 TISSUE TYPE: not applicable
47 CELL TYPE: not applicable
48 ORGANELL: not applicable
49 IMMEDIATE SOURCE: mycobacteriophage L5 particles
50 POSITION IN GENOME: entire genome
51 FEATURE:
52 NAME/KEY:
53 LOCATION:
54 IDENTIFICATION METHOD:
55 OTHER INFORMATION:
56 PUBLICATION INFORMATION:
57 AUTHORS: Hatfull and Sarkis
58 TITLE: DNA Sequence, Structure and Gene
59 TITLE: Expression of Mycobacteriophage L5:
60 TITLE: A Phase System for Mycobacterial
61 TITLE: Genetics.
62 JOURNAL: Molecular Microbiology
63 VOLUME: 7
64 PAGES: 395-405
65 DATE: 1993
66
67 US-08-705-557-1
68
69 Query Match 3.2% Score 34.6; DB 4; Length 52297;
70 Best Local Similarity 49.2% Pctd. No. 2.9;
71 Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0
72
73 431 TTGCGCTTAGACGACCAACCAAGAACCCATTGACGCTTGCGGATGCAACTATGCGTGTATC 4300
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```



LOCATION: 1..1077  
US-08-809-103B-7

Query Match  
Best Local Similarity 50.3%; Pred. No. 4;  
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 856 CTCGTCATTGTAATTCATTCACAGAAACATGTGCACCCCGACCGAAATTTAGAC 915  
DB 172 CTCATATTTCTATTCATCCAAATTCGAGAAATATGATGACCAATCAACGATTTCTGCAC 231

QY 916 ACGGTCTGCAAGACGAGCAAGAAACGTTCCCAACTAGAAAGGAAATCGCAAAATTTGA 975  
DB 232 CTGGTATTCCTCCCAACGAGTCAGACATTTCCATTCGCAACATTCAGGAGCTAAATCGAGC 291

QY 976 CCCATGTTGCTGCGCAAAATTAACAAAGGCGCTGA 1010  
DB 292 TCCGACGTCAAGTCTATATCGACAAGACGAGAGA 326

RESULT 14  
US-08-809-103B-1  
Sequence 1, Application US/08809103B  
Patent No. 6133505

GENERAL INFORMATION:  
APPLICANT: GRONENBORN, Bruno  
TITLE OF INVENTION: PHYTOPATHOGENIC DNA VIRUS RESISTANT  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/809,103B  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94.11040  
FILING DATE: 15-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR95/01192  
FILING DATE: 15-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: US94AL CNR TOM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-2297  
TELEFAX: (703) 685-0573  
TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1148 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1077  
US-08-809-103B-1

Query Match  
Best Local Similarity 50.3%; Pred. No. 4;  
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 856 CTCGTCATTGTAATTCATTCACAGAAACATGTGCACCCCGACCGAAATTTAGAC 915  
DB 172 CTCATATTTCTATTCATCCAAATTCGAGAAATATGATGACCAATCAACGATTTCTGCAC 231

QY 916 ACGGTCTGCAAGACGAGCAAGAAACGTTCCCAACTAGAAAGGAAATCGCAAAATTTGA 975  
DB 232 CTGGTATTCCTCCCAACGAGTCAGACATTTCCATTCGCAACATTCAGGAGCTAAATCGAGC 291

QY 976 CCCATGTTGCTGCGCAAAATTAACAAAGGCGCTGA 1010  
DB 292 TCCGACGTCAAGTCTATATCGACAAGACGAGAGA 326

RESULT 15  
US-08-809-103B-3  
Sequence 3, Application US/08809103B  
Patent No. 6133505

GENERAL INFORMATION:  
APPLICANT: GRONENBORN, Bruno  
TITLE OF INVENTION: PHYTOPATHOGENIC DNA VIRUS RESISTANT  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/809,103B  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94.11040  
FILING DATE: 15-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR95/01192  
FILING DATE: 15-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: US94AL CNR TOM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-2297  
TELEFAX: (703) 685-0573  
TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1150 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1077  
US-08-809-103B-3

Query Match  
Best Local Similarity 50.3%; Pred. No. 4;  
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 856 CTCGTCATTGTAATTCATTCACAGAAACATGTGCACCCCGACCGAAATTTAGAC 915  
DB 172 CTCATATTTCTATTCATCCAAATTCGAGAAATATGATGACCAATCAACGATTTCTGCAC 231





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 1, 2003, 15:32:09 ; Search time 226 Seconds  
(Without alignments)  
7152.939 Million cell updates/sec

Title: US-08-468-190c-1\_COPY\_9874\_10962  
Sequence: 1 ATGCATCGCTACCTGACAC.....CCGACGCAATAAATGGA 1089

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA: \*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/1/pubpna/PC7\_NEM\_PUB.seq: \*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq: \*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq: \*  
6: /cgn2\_6/ptodata/1/pubpna/PCUS\_PUBCOMB.seq: \*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq: \*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq: \*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq: \*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq: \*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq: \*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description         |
|------------|--------|-------------|--------|----|---------------------|
| 1          | 1089   | 100.0       | 13473  | 9  | US-09-994-064-1     |
| 2          | 1087.4 | 99.9        | 3605   | 10 | US-09-881-457A-3    |
| 3          | 1087.4 | 99.9        | 18913  | 9  | US-09-994-064-59    |
| 4          | 346    | 31.8        | 534    | 9  | US-09-994-064-18    |
| 5          | 346    | 31.3        | 1049   | 9  | US-10-123-155-358   |
| 6          | 35     | 3.2         | 653    | 9  | US-10-184-634-402   |
| 7          | 35     | 3.2         | 653    | 9  | US-10-184-634-402   |
| 8          | 34.6   | 3.2         | 475    | 9  | US-10-040-739-899   |
| 9          | 34.6   | 3.1         | 493    | 9  | US-09-918-995-21276 |
| 10         | 34.2   | 3.1         | 802    | 9  | US-10-184-644-312   |
| 11         | 34.2   | 3.1         | 802    | 9  | US-10-184-634-312   |
| 12         | 34     | 3.1         | 2855   | 9  | US-10-125-540-594   |
| 13         | 34     | 3.1         | 2855   | 10 | US-09-764-870-594   |
| 14         | 33.2   | 3.0         | 777    | 9  | US-10-022-832-1     |
| 15         | 32.4   | 3.0         | 491    | 10 | US-09-664-761-6494  |
| 16         | 32.4   | 3.0         | 869    | 9  | US-10-157-031-164   |
| 17         | 32.4   | 3.0         | 3024   | 10 | US-09-833-381-889   |
| 18         | 31.6   | 2.9         | 1223   | 10 | US-09-815-242-9609  |
| 19         | 31.6   | 2.9         | 23105  | 9  | US-10-114-170-39    |

|      |      |     |       |    |                     |                    |
|------|------|-----|-------|----|---------------------|--------------------|
| C 20 | 31.4 | 2.9 | 941   | 9  | US-10-123-155-464   | Sequence 464, App  |
| C 21 | 31.2 | 2.9 | 458   | 9  | US-10-184-644-546   | Sequence 546, App  |
| C 22 | 31.2 | 2.9 | 458   | 9  | US-10-184-634-546   | Sequence 546, App  |
| C 23 | 31.2 | 2.9 | 4166  | 10 | US-09-782-980-73    | Sequence 73, App1  |
| C 24 | 31.2 | 2.9 | 88421 | 9  | US-09-976-059-1     | Sequence 1, App1   |
| C 25 | 31   | 2.8 | 730   | 9  | US-10-226-489-5     | Sequence 5, App1   |
| C 26 | 31   | 2.8 | 1107  | 10 | US-09-974-300-1003  | Sequence 1003, App |
| C 27 | 31   | 2.8 | 1944  | 9  | US-09-943-123-18    | Sequence 18, App1  |
| C 28 | 31   | 2.8 | 1944  | 10 | US-09-917-800A-1735 | Sequence 1735, App |
| C 29 | 30.8 | 2.8 | 1086  | 10 | US-10-123-155-296   | Sequence 296, App  |
| C 30 | 30.8 | 2.8 | 1111  | 9  | US-09-983-446A-1    | Sequence 6138, App |
| C 31 | 30.8 | 2.8 | 2322  | 9  | US-09-983-446A-1    | Sequence 2, App1   |
| C 32 | 30.6 | 2.8 | 335   | 9  | US-10-184-644-174   | Sequence 174, App  |
| C 33 | 30.6 | 2.8 | 335   | 9  | US-10-184-634-174   | Sequence 174, App  |
| C 34 | 30.6 | 2.8 | 405   | 9  | US-09-918-995-3924  | Sequence 3924, App |
| C 35 | 30.6 | 2.8 | 440   | 9  | US-10-184-644-442   | Sequence 442, App  |
| C 36 | 30.6 | 2.8 | 440   | 9  | US-10-184-634-442   | Sequence 442, App  |
| C 37 | 30.6 | 2.8 | 671   | 9  | US-10-184-634-346   | Sequence 346, App  |
| C 38 | 30.6 | 2.8 | 671   | 9  | US-10-184-634-346   | Sequence 346, App  |
| C 39 | 30.6 | 2.8 | 1024  | 9  | US-10-123-155-198   | Sequence 198, App  |
| C 40 | 30.6 | 2.8 | 1323  | 10 | US-09-976-740-45    | Sequence 4076, App |
| C 41 | 30.6 | 2.8 | 1614  | 9  | US-10-023-529-45    | Sequence 45, App1  |
| C 42 | 30.6 | 2.8 | 1614  | 12 | US-10-023-529-45    | Sequence 45, App1  |
| C 43 | 30.6 | 2.8 | 1614  | 12 | US-10-023-529-45    | Sequence 45, App1  |
| C 44 | 30.6 | 2.8 | 2051  | 10 | US-09-917-800A-1325 | Sequence 1325, App |
| C 45 | 30.6 | 2.8 | 2051  | 10 | US-09-917-800A-1325 | Sequence 1325, App |

#### ALIGNMENTS

RESULT 1  
US-09-994-064-1  
Sequence 1, Application US/09994064  
Publication No. US20030082788A1  
GENERAL INFORMATION:  
APPLICANT: Wild, Martha A.  
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/994,064  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/468,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 39116-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 291-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13473 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1059..2489  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2575..4107  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 4113..4445  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 4609..5487  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5697..8654  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9874..10962  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 11159..12658  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 12665..13447  
US-09-994-064-1

Query Match 100.0%; Score 1089; DB 9; Length 13473;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1089; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCAGTCTGCTGGAATCTGCTCTCTCTGCGCGGAGCGCTGCGACCCCTTGGCGCG 60  
DB 9874 ATGGCAGTCTGCTGGAATCTGCTCTCTCTGCGCGGAGCGCTGCGACCCCTTGGCGCG 9933  
QY 61 ATGGGAATCGTGATCTGGAATCTGCTCTCTCTGCGCGGAGCGCTGCGACCATCTGCTG 120  
DB 9934 ATGGGAATCGTGATCTGGAATCTGCTCTCTCTGCGCGGAGCGCTGCGACCATCTGCTG 9993  
QY 121 ATGCTGCGCGCTGCGCGCGGAGCTGAGTCAATTCAGTCACTGAGTATTTTCATGCTGGCGCG 180  
DB 9994 ATGCTGCGCGCTGCGCGCGGAGCTGAGTCAATTCAGTCACTGAGTATTTTCATGCTGGCGCG 10053  
QY 181 AGACCCCAAAACCTCTAGAGAACCGCTGCGCGCTGCTGCTGATATTAACAAC 240  
DB 10054 AGACCCCAAAACCTCTAGAGAACCGCTGCGCGCTGCTGCTGATATTAACAAC 10113  
QY 241 CAGTGTCTACGAGAACTTACGAGAGCGCTTTGAAAATTCGACTCATGATCGTCTCT 300  
DB 10114 CAGTGTCTACGAGAACTTACGAGAGCGCTTTGAAAATTCGACTCATGATCGTCTCT 10173  
QY 301 GTTTTGTGCGCTGTAAGTAGCGAGTACAGTCTCTGCGCTGGAACAGTAAACCGGA 360  
DB 10174 GTTTTGTGCGCTGTAAGTAGCGAGTACAGTCTCTGCGCTGGAACAGTAAACCGGA 10233  
QY 361 CCTCCACACCGCTTAAAGCTACATATAGAAATCTCTGCTCGGAAGAGCGGGATGTT 420  
DB 10234 CCTCCACACCGCTTAAAGCTACATATAGAAATCTCTGCTCGGAAGAGCGGGATGTT 10293  
QY 421 TACGTATTTGCTGCGCTAGAGAGACCAACCAATTCAGTCTGCGATCAACTA 480  
DB 10294 TACGTATTTGCTGCGCTAGAGAGACCAACCAATTCAGTCTGCGATCAACTA 10353  
QY 481 TCGGTATCAATTCGGAACACCGCGAGCTGCGGAGCTATTTCCAAAGGCTTGGT 540  
DB 10354 TCGGTATCAATTCGGAACACCGCGAGCTGCGGAGCTATTTCCAAAGGCTTGGT 10413  
QY 541 CGCAGCTTGGATTACCTACCGTCAACTTGAAGCTATTCAGAGCGAGAAAGTTGG 600  
DB 10414 CGCAGCTTGGATTACCTACCGTCAACTTGAAGCTATTCAGAGCGAGAAAGTTGG 10473  
QY 601 CGCAACTGGCAAGCGTGTGCGACGAGGCGCACGAGACCAAGCGCGGAGCGCAAC 660

DB 10474 CGCAACTGGCAAGCGTGTGCGACGAGGCGCACGAGACCAAGCGCGGAGCGCAAC 10533  
QY 661 CCGAGCGCGCTGCTGCTGCAACAGCGCGCTCCCAACTGTAAGGGAACACTTACCTTCC 720  
DB 10534 CCGAGCGCGCTGCTGCTGCAACAGCGCGCTCCCAACTGTAAGGGAACACTTACCTTCC 10593  
QY 721 TGGCTAGAAAATTTGGCGTGTGATCATTTACGACCGACCGCGAAAGAAAATTTCAACGTT 780  
DB 10594 TGGCTAGAAAATTTGGCGTGTGATCATTTACGACCGACCGCGAAAGAAAATTTCAACGTT 10653  
QY 781 ACTGTCGCTGCTGCGGACATGAGCCCTACCTAATTTGGGTTAACCGTGGCTGCTG 840  
DB 10654 ACTGTCGCTGCTGCGGACATGAGCCCTACCTAATTTGGGTTAACCGTGGCTGCTG 10713  
QY 841 AGCGCAAGCATGCGCTGCTGATGTATTTTCATTCGTCACAGAAACATGTGACACCCG 900  
DB 10714 AGCGCAAGCATGCGCTGCTGATGTATTTTCATTCGTCACAGAAACATGTGACACCCG 10773  
QY 901 CACGGAATTTAGACACGCTGCTGCAAGACGAGCAAGAACGTTCCCAACTAGAAAGGAA 960  
DB 10774 CACGGAATTTAGACACGCTGCTGCAAGACGAGCAAGAACGTTCCCAACTAGAAAGGAA 10833  
QY 961 TCGGGAATTTTGGACCCATGCTGCTGCGAATTAACAAGGCGCTGACCGAGATAGT 1020  
DB 10834 TCGGGAATTTTGGACCCATGCTGCTGCGAATTAACAAGGCGCTGACCGAGATAGT 10893  
QY 1021 GAACCTTGGGAAGTGGTGGATTTGTAACCGCTGCTGCGTACGCTGCGCGAGCTCATTA 1080  
DB 10894 GAACCTTGGGAAGTGGTGGATTTGTAACCGCTGCTGCGTACGCTGCGCGAGCTCATTA 10953  
QY 1081 AAAATGTGA 1089  
DB 10954 AAAATGTGA 10962

RESULT 2  
US-09-881-457A-3  
Sequence 3, Application US/09881457A  
Patent No. US20020081316A1  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D  
APPLICANT: Cook, Stephanie M  
APPLICANT: Wild, Martha A  
TITLE OF INVENTION: No. US20020081316A1e1 Avian Herpes Virus and Uses Thereof  
FILE REFERENCE: SY01105K10KOK  
CURRENT APPLICATION NUMBER: US/09/881,457A  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: 09/426,352  
PRIOR FILING DATE: 1999-10-25  
PRIOR APPLICATION NUMBER: 08/804,372  
PRIOR FILING DATE: 1997-02-21  
PRIOR APPLICATION NUMBER: PCT/US95/10245  
PRIOR FILING DATE: 1995-08-09  
PRIOR APPLICATION NUMBER: 08/663,566  
PRIOR FILING DATE: 1996-06-13  
PRIOR APPLICATION NUMBER: 08/288,065  
PRIOR FILING DATE: 1994-08-09  
PRIOR APPLICATION NUMBER: PCT/US93/05681  
PRIOR FILING DATE: 1993-06-14  
PRIOR APPLICATION NUMBER: 08/023,610  
PRIOR FILING DATE: 1993-02-26  
PRIOR APPLICATION NUMBER: 07/898,087  
PRIOR FILING DATE: 1992-06-12  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 3605  
TYPE: DNA  
ORGANISM: Infectious Latyngotracheitis Virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (585)..(1889)

OTHER INFORMATION: ILTV glycoprotein D  
NAME/KEY: CDS  
LOCATION: (1997)..(3085)  
OTHER INFORMATION: ILTV glycoprotein I  
US-09-881-457A-3

Query Match 99.9% Score 1087.4 DB 10; Length 3605;  
Best Local Similarity 99.9% Freq. No. 0;  
Matches 1088; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCATGCTACTGTAAGTCTGCTCTCTCTCCGCGAGCTGCGACCCCTTGCGCGC 60  
DB ATGGCATGCTACTGTAAGTCTGCTCTCTCTCTCCGCGAGCTGCGACCCCTTGCGCGC 2056  
QY 1997 ATGGCATGCTACTGTAAGTCTGCTCTCTCTCTCCGCGAGCTGCGACCCCTTGCGCGC 2056  
DB 61 ATGGCAATGCTATGTAAGTCTGCTCTCTCTCTCCGCGAGCTGCGACCCCTTGCGCGC 120  
QY 2057 ATGGCAATGCTATGTAAGTCTGCTCTCTCTCTCCGCGAGCTGCGACCCCTTGCGCGC 2116  
DB 121 ATGCTGCGCCCTGCGCCCGAGAGTCAATTAATCACTGCACTATTTTCATGCTGCGCAG 180  
QY 2117 ATGCTGCGCCCTGCGCCCGAGAGTCAATTAATCACTGCACTATTTTCATGCTGCGCAG 2176  
DB 181 AGACCCCAAAACCTACTAGAAACGCTCGGCTGCTGCTGATATTAACAAAC 240  
QY 2177 AGACCCCAAAACCTACTAGAAACGCTCGGCTGCTGCTGATATTAACAAAC 2236  
DB 241 CAGTGTACAGAGACTAG 300  
QY 2237 CAGTGTACAGAGACTAG 2296  
DB 301 GTTTTGTGCGCTGTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 360  
QY 2297 GTTTTGTGCGCTGTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2356  
DB 361 CTTCCACACCGCTTAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 420  
QY 2357 CTTCCACACCGCTTAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2416  
DB 421 TACGTAATGTTGCGCTAGAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 480  
QY 2417 TACGTAATGTTGCGCTAGAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 2476  
DB 481 TCGGTATATCAATTTGCGACACAGCGCGCGACCTGATTCATTCAGAGCTTCGTGT 540  
QY 2477 TCGGTATATCAATTTGCGACACAGCGCGCGACCTGATTCATTCAGAGCTTCGTGT 2536  
DB 541 CGACCTTGCATTTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 600  
QY 2537 CGACCTTGCATTTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2596  
DB 601 CGACCTTGCATTTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 660  
QY 2597 CGACCTTGCATTTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2656  
DB 661 CGACCTTGCATTTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 720  
QY 2657 CGACCTTGCATTTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2716  
DB 721 TGGCTAGAAAATGGCGTGAATCAAGACAGACAGACAGACAGACAGACAGACAGACAG 780  
QY 2717 TGGCTAGAAAATGGCGTGAATCAAGACAGACAGACAGACAGACAGACAGACAGACAG 2776  
DB 781 ACTGTCCGCTCGGGAGCAATGAGCCCTACCTAATTTGGGGTAAACCGTGGCTGCTG 840  
QY 2777 ACTGTCCGCTCGGGAGCAATGAGCCCTACCTAATTTGGGGTAAACCGTGGCTGCTG 2836  
DB 841 AGGCAACAGATGAGGCTGCTATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTT 900  
QY 2837 AGGCAACAGATGAGGCTGCTATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTT 2896  
DB 901 CACCGAAAATTTAGACAGAGTCTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 2897 CACCGAAAATTTAGACAGAGTCTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2956  
DB

QY 961 TCGGAAAATTTGGACCATGGTGGTGGCCGAATTAACAAGGGCGCTGACCGAGATAGT 1020  
DB 2957 TCGGAAAATTTGGACCATGGTGGTGGCCGAATTAACAAGGGCGCTGACCGAGATAGT 3016  
QY 1021 GACTGTGGAACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080  
DB 3017 GACTGTGGAACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3076  
QY 1081 AAAATGTGA 1089  
DB 3077 AAAATGTGA 3085

RESULT 3  
US-09-994-064-59  
Sequence 59, Application US/09994064  
Publication No. US20030082788A1  
GENERAL INFORMATION:  
APPLICANT: Wild, Martha A.  
TITLE OF INVENTION: RECOMBINANT INFECTIOUS IARIYNGOTRACHEITIS VIRUS  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/994,064  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/468,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 39116-A  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18913 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 697..1533  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (1900..2784)  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (2916..3605)  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: CDS

```
LOCATION: 3694..5124
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 5210..7081
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 7245..8123
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 8333..11290
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 11098..12402
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 12510..13598
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 13792..15291
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 15298..16080
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 16129..17013
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: 17380..18216
OTHER INFORMATION:
US-09-994-064-59

Query Match          99.9%: Score 1087.4; DB 9; Length 18913;
Best Local Similarity 99.9%: Pred. No. 0;
Matches 1088; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGGCATGCTACTTGGAACTGTGCTCTCTGCGCGAGCGCTCGACCCCTTGCGGCG 60
DB 12511 ATGGCATGCTACTTGGAACTGTGCTCTCTGCGCGAGCGCTCGACCCCTTGCGGCG 12570
DB 61 ATGGGAATCGTGATCACTGGAATCAAGTCTCGCGCGAGATTGACGACGATCACATCGTG 120
DB 12571 ATGGGAATCGTGATCACTGGAATCAAGTCTCGCGCGAGATTGACGACGATCACATCGTG 12630
OY 121 ATGTCGCGCGCTGCGCGCGAGAGCTACAAATCACTGAGCTAATTTTTCATGCGCGCAG 180
DB 12631 ATGTCGCGCGCTGCGCGCGAGAGCTACAAATCACTGAGCTAATTTTTCATGCGCGCAG 12690
OY 181 AGACCCCAAAACCTCTACTGAGAAACCGTCCGCGTGGCTGTGATATAACAAC 240
DB 12691 AGACCCCAAAACCTCTACTGAGAAACCGTCCGCGTGGCTGTGATATAACAAC 12750
OY 241 CAGTGTACAGAGAACTTACGAGAGCGCTTTGAAAATTGCACTCTCATGCTCTTCT 300
DB 12751 CAGTGTACAGAGAACTTACGAGAGCGCTTTGAAAATTGCACTCTCATGCTCTTCT 12810
OY 301 GTTTTGTGCGGTGTAAAGTACGAGAGTACAGCTTCTCGCGTGAAGACAGTAAACGGA 360
DB 12811 GTTTTGTGCGGTGTAAAGTACGAGAGTACAGCTTCTCGCGTGAAGACAGTAAACGGA 12870
OY 361 CCTGCACACCCGTTTAACTCACTATACGAAATCTCTGTCGGAAGCAGACGGGATGTC 420
DB 12871 CCTGCACACCCGTTTAACTCACTATACGAAATCTCTGTCGGAAGCAGACGGGATGTC 12930
OY 421 TACGTAATTTGTGGGCTAGACGACACCAAGAACCCATTGACGCTTTCGGAATCAACTA 480
```

```
|||||
DB 12931 TACGTAATTTGTGGGCTAGACGACACCAAGAACCCATTGACGCTTTCGGAATCAACTA 12990
OY 481 TCGGTATATCAATTCGGGAACACCGCGCGAGCTGCGGAGCTTTCGAAGGCTTCGTGT 540
DB 12991 TCGGTATATCAATTCGGGAACACCGCGCGAGCTGCGGAGCTTTCGAAGGCTTCGTGT 13050
OY 541 CGCACTTTCGATTAATCTTACCGTCCAACTTGAGGCTTATCTCAGAGCCGAGAAAGTTGG 600
DB 13051 CGCACTTTCGATTAATCTTACCGTCCAACTTGAGGCTTATCTCAGAGCCGAGAAAGTTGG 13110
OY 601 CGCACTTTCGATTAATCTTACCGTCCAACTTGAGGCTTATCTCAGAGCCGAGAAAGTTGG 660
DB 13111 CGCACTTTCGATTAATCTTACCGTCCAACTTGAGGCTTATCTCAGAGCCGAGAAAGTTGG 13170
OY 661 CGCACTTTCGATTAATCTTACCGTCCAACTTGAGGCTTATCTCAGAGCCGAGAAAGTTGG 720
DB 13171 CGCACTTTCGATTAATCTTACCGTCCAACTTGAGGCTTATCTCAGAGCCGAGAAAGTTGG 13230
OY 721 TGGCTAGAAATGCGTGTGATCATTAAGAAACGACACCCGCAAGAAATTCAAAGCTT 780
DB 13231 TGGCTAGAAATGCGTGTGATCATTAAGAAACGACACCCGCAAGAAATTCAAAGCTT 13290
OY 781 ACTGTCCGTCTCGGGAACATGAGCCCTACGCTAATTTGGGTACCGTGGCTGCCGTGTG 840
DB 13291 ACTGTCCGTCTCGGGAACATGAGCCCTACGCTAATTTGGGTACCGTGGCTGCCGTGTG 13350
OY 841 AGCGCAAGATGCGGCTGCTGATGTATTTTCCATGCTCAACGAAATGTGCAACCCG 900
DB 13351 AGCGCAAGATGCGGCTGCTGATGTATTTTCCATGCTCAACGAAATGTGCAACCCG 13410
OY 901 CACGGAATTTAGACACGCTGCTGCAAGACGACGAGACGATTCACAACTAGAAAGGAA 960
DB 13411 CACGGAATTTAGACACGCTGCTGCAAGACGACGAGACGATTCACAACTAGAAAGGAA 13470
OY 961 TCGGGAATTTTGGACCAATGCTTGGCTGCGGAATTAACAAGGCGCTGACAGATAGT 1020
DB 13471 TCGGGAATTTTGGACCAATGCTTGGCTGCGGAATTAACAAGGCGCTGACAGATAGT 13530
OY 1021 GAACCTTGGAACTGTTGATGATTTTAAACCCGCTGCGTAAGCTGCGGAGTCAATA 1080
DB 13531 GAACCTTGGAACTGTTGATGATTTTAAACCCGCTGCGTAAGCTGCGGAGTCAATA 13590
OY 1081 AAAATGTGA 1089
DB 13591 AAAATGTGA 13599

RESULT 4
US-09-994-064-18/c
Sequence 18, Application US/09994064
Publication No. US20030082788A1
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/994,064
FILING DATE:
```

```

CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/468,190
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39116-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..533
US-09-994-064-18

Query Match
Best Local Similarity 31.8%; Score 346; DB 9; Length 534;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 TTACGAACCGACACCGCGCAACGAAATTTCAACGTTACTGTCCTGCTGGGACAAATGAG 803
DB 534 TTACGAACCGACACCGCGCAACGAAATTTCAACGTTACTGTCCTGCTGGGACAAATGAG 475
QY 804 CCTACGCTAATTTGGGGTAACCGTGGCTGCGTGTAGGCGCAACGATGGGCTGTGAT 863
DB 474 CCTACGCTAATTTGGGGTAACCGTGGCTGCGTGTAGGCGCAACGATGGGCTGTGAT 415
QY 864 TGTAAATTCATGCTGACAGAAACATGTGACCCCGCCAGCGAAATTTAGACACGCTCTC 923
DB 414 TGTAAATTCATGCTGACAGAAACATGTGACCCCGCCAGCGAAATTTAGACACGCTCTC 355
QY 924 GCAAGACGACGAAAGACGTTCCCAACTAGAAAGGAAATCGGAAATTTGGACCCATGGT 983
DB 354 GCAAGACGACGAAAGACGTTCCCAACTAGAAAGGAAATCGGAAATTTGGACCCATGGT 295
QY 984 TGGCGGGAATTAACAAGAGGGGCTGACAGGATAGTGAACCTGTGGAACGCTGTGGAT 1043
DB 294 TGGCGGGAATTAACAAGAGGGGCTGACAGGATAGTGAACCTGTGTGAACCTGTGGAT 235
QY 1044 TGTAAACCGCTGCGCTAAGCTGCGCCGACTCAATATAAATGTGA 1089
DB 234 TGTAAACCGCTGCGCTAAGCTGCGCCGACTCAATATAAATGTGA 189

RESULT 5
US-10-123-155-358/c
Sequence 358, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin J.
APPLICANT: Godowski, Paul J.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel

```

```

APPLICANT: Matanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C30
CURRENT APPLICATION NUMBER: US/10/123,155
PRIOR APPLICATION DATE: 2002-04-15
Pilot Application removed - see Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 358
LENGTH: 1049
TYPE: PRT
ORGANISM: Homo Sapien
US-10-123-155-358

Query Match
Best Local Similarity 3.3%; Score 36; DB 9; Length 1049;
Matches 35; Conservative 215; Mismatches 529; Indels 1; Gaps 1;

QY 270 CTTTGAAATTCGACTCATGATGCTCTCTTTTGTTCGGCTGTAAAGTGAACGAGTA 329
DB 818 C.CBGBHAT.R..AB.ADNBMB.WETHAG..YMD..NM..NN..BN..SAM..K.. 759
QY 330 CAGCTTCGCTCGAAGACAGTAACCGACCTCCACCCGTTAAGTCACTATACG 389
DB 758 MNSS..H.RY..TH..RMA..SY..NNM..NN.SYSGNS.Y..BAA..NDS..H.A..NM 699
QY 390 AATTCCTC-GTCCGAACGACAGCGGATGTTCTACGTAATTTGCTGAGACACCA 448
DB 698 .G..MMMS..SM..CNTT..S.NM.N..KCH..BCS..S.SNMS..H..M..NM..RYNH 639
QY 449 AAGAACCCATTCACCTCTTCGCGATCCCAATATGATGATCAATTCGCAACCGCGG 508
DB 638 C.YW..BH..DNCY..AY..S.S.KAVSASS.HHNHKK..M..BM..NMA..N.KDA..C.S..R 579
QY 509 CGACTGCGGACTATATTCACGAGCTTCGCTGTCGACCTTCGATTAAGTCACTCAAC 568
DB 578 DSNSS..H.B..MD...TASD..H.YNNS..H.RY..T...SCN..A.S..NCS..N.GM.. 519
QY 569 TTGAGGCTATCTGACGACGAGGAAAGTGGCGCAATGCGCAACGATGTCGACG 628
DB 518 .S.D..HSSMB...SNMS..H.A..CRMGS..NBSK..ST..NMN..YGYTRHRY..RD...B 459
QY 629 AGGCCACGACGACGACGCGCGGAGCGCAACCGCGCGCTGCTGCAACGACGCGCT 688
DB 458 ..RS..BSATVNSG..CB..SSHCS..S.MNBS..H.BM..YM..M.KS..NT..M..MAC..H.B.. 399
QY 689 CCGAAGCTTAAGCGAAGACTTTTACCTTTCCCTGCTAGAAATTTGCGTGCATTAAG 748
DB 398 N..ND..S..N..SM..M..BRCY..Y..M..SM..SS..T..S..NKSTYRB...N..S.H...S.. 339
QY 749 AACCGACACCGCAACGAAATTAACAAGCTTACGCTGCTGCGGACAAATGACCCCTA 808
DB 338 .D..MTHC..MT..N.S..H...MN..NM..WT..BD..SNSD..Y..BM..A..TH..TNB... 279
QY 809 CGCTAATTTGGGGTAACCGTGGCTCCCTGCTGAGCGCAACGATGCGCTGCTGATGTA 868
DB 278 .SNMG..TG...TNRGY..GNCS..H...N..NN..HH...MT..KNNR..R..A..AS..BA..B 219
QY 869 TTTCATGCTCACGAAACATGTGACCCCGGACGGAATTAAGACACGCTCCGCAAG 928
DB 218 TABNNHM..S..BW..MA..N..THM..SREBRG..NYRRON..C.R...NT..A..N..MY..S..N 159
QY 929 ACAGCAAGAAAGTTCCCAACTAGAAAGGAATCGGAAATTTGGAACCCATGTTGCGT 988
DB 158 NT..S...S...C.....NCH..R..SM..RA..CS..SY..M...YM..GKNMNC...BCN 99
QY 989 GCGAAATTAACAAGGCGCGGACAGATGTCGTAACCTGTGACGCTGTCGATTCGTA 1048
DB 98 GY..H..B..DH..YD..ST..S..H..DN..A..A..NAANA..CC...A..DHHAGHB..BDNM..BH..A 39

```

```
RESULT 6
US-10-184-644-402/c
; Sequence 402, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC227
; CURRENT APPLICATION NUMBER: US/10/184,644
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-402

Query Match
Best Local Similarity 5.0%; Pired. No. 0.2;
Matches 24; Conservative 160; Mismatches 298; Indels 0; Gaps 0;

14 TTGGAACCTGCTCCTCTCCCTGCGGACGCTCGACCCCTTCGCGCGATGGAAATCGTA 73
518 YSSYS.S.SYNMMDMDYDCSSSSSSSSBS.HHSBSSSSASMYMM.YM.M.MY 459
74 TCACGTGAATACGCTCCGCGGAGATTGACGACGATCATCGTGCCTGCGGCTC 133
458 .M.TSA..MMAM.M.M.T.M.Y...SM.KRMD.M.TH.THM.A.H.T.TMRRS.TN. 399
134 GCGCCGAGCTCAATTCACGACGCTATTTTCATGCTGCGGAGAGACCCACCAAC 193
398 .M.....CCY.B.A.G.RMYTNYDA.GN.T...H.TMN.SCMATR.TCYTB.T.BN. 339
194 CCTACTGAGACCGCTCCGCGGCTTCGCTGATATACAAACGAGCTCTACGAG 253
338 MH...TMNR.NKTHDYCBM.RHBC.MBGM.TWSTS.MMY.TST.HH.S.NMS..CYK.S 279
254 AACTTAGCAGAGCGCTTGAATAATGCATCATGATCGTCTCTGTTTGTGCGCT 313
278 .NS.H.C.M....BBC.NB.C.SS.KBN.RS..SNSN..BSYR...SS.BC.M. 219
314 GTAAGTGACGAGTACAGCTTCGCGCTCGACGACTAACGAGCTCCACACCGCT 373
218 CS.D...SSSR.SBT.M.DRYH.HM.CTY..HCA.RRS..HCDNSDS.BHY..G.TA... 159
374 TTAAGCTCACTATAGAAATCCTCGCGCAAGACGAGCGGATGTCTAGTAATGTTC 433
158 DT.HRK.CSC.G..BK..C.Y.SS.YC.B.HCY..HY...YHKS...K.....K.T 99
434 GGCTAGACGACCAAGAACCATTTGACGCTTCGCGATCCAACTATGCGTATCAAT 493
98 RDH.SH.N..SAT.THSWMS.T...HTWMS..M...NBYM.MYC...D...CYT..T 39
494 TC 495
1:
38 TS 37
```

```
RESULT 7
US-10-184-634-402/c
; Sequence 402, Application US/10184634
```

```
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-402

Query Match
Best Local Similarity 5.0%; Pired. No. 0.2;
Matches 24; Conservative 160; Mismatches 298; Indels 0; Gaps 0;

14 TTGGAACCTGCTCCTCTCCCTGCGGACGCTCGACCCCTTCGCGCGATGGAAATCGTA 73
518 YSSYS.S.SYNMMDMDYDCSSSSSSSSBS.HHSBSSSSASMYMM.YM.M.MY 459
74 TCACGTGAATACGCTCCGCGGAGATTGACGACGATCATCGTGCCTGCGGCTC 133
458 .M.TSA..MMAM.M.M.T.M.Y...SM.KRMD.M.TH.THM.A.H.T.TMRRS.TN. 399
134 GCGCCGAGCTCAATTCACGACGCTATTTTCATGCTGCGGAGAGACCCACCAAC 193
398 .M.....CCY.B.A.G.RMYTNYDA.GN.T...H.TMN.SCMATR.TCYTB.T.BN. 339
194 CCTACTGAGACCGCTCCGCGGCTTCGCTGATATACAAACGAGCTCTACGAG 253
338 MH...TMNR.NKTHDYCBM.RHBC.MBGM.TWSTS.MMY.TST.HH.S.NMS..CYK.S 279
254 AACTTAGCAGAGCGCTTGAATAATGCATCATGATCGTCTCTGTTTGTGCGCT 313
278 .NS.H.C.M....BBC.NB.C.SS.KBN.RS..SNSN..BSYR...SS.BC.M. 219
314 GTAAGTGACGAGTACAGCTTCGCGCTCGACGACTAACGAGCTCCACACCGCT 373
218 CS.D...SSSR.SBT.M.DRYH.HM.CTY..HCA.RRS..HCDNSDS.BHY..G.TA... 159
374 TTAAGCTCACTATAGAAATCCTCGCGCAAGACGAGCGGATGTCTAGTAATGTTC 433
158 DT.HRK.CSC.G..BK..C.Y.SS.YC.B.HCY..HY...YHKS...K.....K.T 99
434 GGCTAGACGACCAAGAACCATTTGACGCTTCGCGATCCAACTATGCGTATCAAT 493
98 RDH.SH.N..SAT.THSWMS.T...HTWMS..M...NBYM.MYC...D...CYT..T 39
494 TC 495
1:
38 TS 37
```

```
RESULT 8
US-10-040-739-899
; Sequence 899, Application US/10040739
; Patent No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
```

McCoY, John  
Lavalie, Edward  
Racie, Lisa  
Merberg, David  
Treacy, Maurice  
Spaulding, Vikki  
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS  
NUMBER OF SEQUENCES: 1519  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/040,739  
FILING DATE: 07-Jan-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/036,520  
FILING DATE: 03-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5951

INFORMATION FOR SEQ ID NO: 899:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 475 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 899:

US-10-040-739-899

Query Match 3.2%; Score 34.6; DB 9; Length 475;  
Best Local Similarity 50.9%; Pred. No. 0.24;  
Matches 82; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

154 CTGCAGTATTTTCATGCGCTGCCAGACAGACCCACAAACCCTACTCAGACGCTCCG 213  
+ + + + +  
239 CTGTAACTAGTTGGCTCGCAGGCGCTGTTGTTCAAGTAGAGATTGTGCCCCAACAA 298  
+ + + + +  
OY 214 CTCGCCGTTTGCGCTCATATTAACAACAACAGCTGCTACCAAGAACTAGCAGGAGAGCCITT 273  
+ + + + +  
Db 239 ATGGGGTGGTGGGCTCTGCTCCACCAAATCTTTATCCAGTCTTCACTCGAAGAGAGGCGACT 358  
+ + + + +  
OY 274 GAATAATGCACATCATGATCGCTCTCTGTTGTTTTGTGGGGTG 314  
+ + + + +  
Db 359 GCCAAGTGTCTCAGCCAGACAGTCTTGTGTCTAAGGCTCTG 399  
+ + + + +

RESULT 9  
US-09-918-995-21276/c  
Sequence 21276, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT FILING DATE: 2001-07-30  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20

```

: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 21276
: LENGTH: 493
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: (1)...(493)
: OTHER INFORMATION: n - A,T,C or G
US-09-918-995-21276

Query Match
Best Local Similarity 3.28; Score 34.6; DB 9; Length 493;
Matches 82; Conservative 0; Mismatches 79; Indels 0; Gaps 0

OY 154 CTGACGATATTTTTCATGCTTGCCGACAGACCCACACAACTTACTGAGAACCGTCGCC 213
Db 474 CTGTACATATGTTCTCGGAGGCGCTGTTGTTCAGAGTAGAGCTTGCGCCAGACAA 415
OY 214 GTGCGGTTTGCGCTCATTTAAACAACCAAGTGTACAGAACTTAGAGGAGCGCTTT 273
Db 414 ATGGGCGTGGGCGCTGCTGCCACAAATCTTTATCCAGTTCATCCAGAGGGGCGACT 355
OY 274 GAAATTCACATCATGATCGATCGTCTCTTTTGTGCGCTG 314
Db 354 GCCAAGTGTCTTCAGCCAGAGTCTTGTGAAGGCTCG 314

RESULT 10
US-10-184-644-312/C
: Sequence 312 Application US/10184644
: Publication No. US20030044930A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Destroyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zhenli
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3430R1C227
: CURRENT APPLICATION NUMBER: US/10/184.644
: CURRENT FILING DATE: 2002-06-28
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 312
: LENGTH: 802
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-184-644-312

Query Match
Best Local Similarity 3.18; Score 34.2; DB 9; Length 802;
Matches 46; Conservative 215; Mismatches 467; Indels 0; Gaps 0

OY 350 GACATACCGGACCTCCACACCCGTTTAAGCTACTATAGCAAAATCTGTCGCCAGACAA 409
Db 762 KKR..M..H..N..HG..B..DDBDBCB..ST..AM..CT..BBMN..HHH..CWC..SBH..CCBYB.. 703
OY 410 GGGGATGTTCTACGTAATTTGTCGGCTAGACGACGACCAAGAACCCATTGAGCTCTGC 469
Db 702 HCMR..GA..C..CRNYW..CA..A..T..DNHS..BMSRB..M..HR..S.....R...C..BANTGY 643
OY 470 GCATCCACTATGCGGTGTATCAATTCGCGAAGCCGCCGCGACTGCGGAGCTATTCCA 529
Db 642 ....AA..B..HBHG.....S..NN..SSCB..T..TYS..CSB.....KTTM..RM..YRHYK..B.. 583

```

```

OY      530  +GGGCTTCGTGGACCGCTTGCATTAACCTACCGCTCACTTGAGCGCTATGCGACGACG 589
Db      582  MTMH.NGNSN...BB.MBN.N...GAM..NCK.YB.KH.YCS...N.MMHM..MD..SM 523
OY      590  AGGAAAGTTGGCGCAACTGCGCAAGCGTACGTTGCCACGAGGAGCGCCAGCAGCAACGACGCGG 649
Db      522  SC....B.MM.SNS..S.SCS..N.YMT...TH...D.B...MS.A...RTDYB..BAK 463
OY      650  AGGGGACAAACCCCGACGCGCGCTCACTGCAACCAACGAGCGCTCGCAACTTGAGCGGAACT 709
Db      462  MMCDMMH...H..R.TCRK.NBYRCR...M.H..YCAVMTTN..K.KB..KH..T.. 403
OY      710  TTACCTTCCCTGGCTAGMAAATGGCGCTGATCATTAACGAACCGACCCGCAAGAA 769
Db      402  Y.TSHKCY...CHBTSR.RMCA...W....Y.Y...YK.S..C...H.MD..ANSRM 343
OY      770  ATTCAAAAGTACTGTCGCTCGTTCGGGACAAATGAGCCCTACGCTAATTGGGTGAACCGTGG 829
Db      342  SK..B..YD...AYDY..S.MYS..KRS.D.YR..R..NMN.D.A.T.D.MEND..HY..RCMN.. 283
OY      830  CTGCGCGTGTGAGCGCAAGATCGGCGCTGTCATGTGAATTTCATGCGTCAACCAAGAA 889
Db      282  ..RN..R...YK..RSMBG..BCT..YVBGYCB..BH..DAAR..Y..GMC.D.BKYY..B.YK..BC. 223
OY      890  TGTGCACCCCGCACCGAAATTTAGACACGCTGTGCGCAAGACGACGAAAGCGTTCCCAA 949
Db      222  CCKG..N.C...T..MCK..AAC..CA.C...SSN..SY..N..YHGM..RBHNTYK..W..RM 163
OY      950  CTAGAAGGAAATCGGAAATTTGGACCCATGCTGCGTGGCAATTAACAAGGCGCTG 1009
Db      162  H.RHDDKRM..K.K.SMM...RSHHBY..BB...BSAHSC..SS...BMC..ANSMAYRTTB 103
OY      1010  ACCAGATAGTGAACCTTGTGGACACTGCTGGCATGTGTAACCCGCTCGCTGAACCTGCG 1065
Db      102  TYA...RM.TAMB..CB...NVHY..CGH..HSC..W...TCYHCTCCG..STTT..CSYGC..ST 43
OY      1070  CCGACTCA 1077
Db      42  YYY.CTYM 35

RESULT 11
US-10-184-634-312/c
: Sequence 312, Application US/10184634
: Publication No. US20030068684a1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430R1C217
: CURRENT APPLICATION NUMBER: US/10/184,634
: PRIOR FILING DATE: 2002-06-28
: Prior Application removed - See file Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 312
: LENGTH: 802
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-184-634-312

```

|                       |       |                 |       |             |
|-----------------------|-------|-----------------|-------|-------------|
| Query Match           | 3.1%; | Score 34.2;     | DB 9; | Length 802; |
| Best Local Similarity | 6.3%; | Pred. No. 0.43; |       |             |

[illegible]

```

RESULT 12 540-594
US-10-125-540-594
: Sequence 594, Application US/70125540
: Publication No. US20030059875A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.,
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: Ptz141
: CURRENT APPLICATION NUMBER: US/10/125,540
: CURRENT FILING DATE: 2002-04-19
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 646
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 594
: LENGTH: 2855
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-125-540-594

```



Query Match 3.1%; Score 34; DB 9; Length 2855;  
Best Local Similarity 57.5%; Pred. No. 0.98;  
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 596 GTTGGCGCACTGCAAGCTGACCTGTCGACGAGCCAGCAGCAGCCGAGGCGA 655  
DB 2122 GGTGCCCCAGCGGCGAGCGGCGGCGGCGGAGCCGAGCCAGCCAGGAGGCT 2181  
QY 656 CAACCCGAGCGCCGCTCAGTCAACCAACCGCCCTCGAAGCTTGAAGC 701  
DB 2182 CCCACCGCGCTCCCTCCGACAGCTGAGAGACATGAAACGAGAACG 2227

## RESULT 13

US-09-764-870-594  
Sequence 594, Application US/09764870  
Patent No. US20020042386A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P214  
CURRENT APPLICATION NUMBER: US/09/764,870  
CURRENT FILING DATE: 2001-01-17  
PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 646  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 594  
LENGTH: 2855  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-870-594

Query Match 3.1%; Score 34; DB 10; Length 2855;  
Best Local Similarity 57.5%; Pred. No. 0.98;  
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 596 GTTGGCGCACTGCAAGCTGACCTGTCGACGAGCCAGCAGCAGCCGAGGCGA 655  
DB 2122 GGTGCCCCAGCGGCGAGCGGCGGCGGCGGAGCCGAGCCAGCCAGGAGCTG 2181  
QY 656 CAACCCGAGCGCCGCTCAGTCAACCAACCGCCCTCGAAGCTTGAAGC 701  
DB 2182 CCCACCGCGCTCCCTCCGACAGCTGAGAGACATGAAACGAGAACG 2227

## RESULT 14

US-10-022-832-1  
Sequence 1, Application US/10022832  
Publication No. US20030039896A1  
GENERAL INFORMATION:

APPLICANT: HAMEL, JOSE  
APPLICANT: BRODEUR, BERNARD R.  
TITLE OF INVENTION: NOVEL CHLAMYDIA ANTIGENS AND CORRESPONDING DNA  
FILE REFERENCE: BIOVAC-15  
CURRENT APPLICATION NUMBER: US/10/022,832  
CURRENT FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/256,941  
PRIOR FILING DATE: 2000-12-21  
NUMBER OF SEQ ID NOS: 77  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 777  
TYPE: DNA  
ORGANISM: Chlamydia pneumoniae  
US-10-022-832-1

Query Match 3.0%; Score 33.2; DB 9; Length 777;  
Best Local Similarity 56.4%; Pred. No. 0.94;  
Matches 62; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 929 ACAGCAAGACCTTCCCAAGTAAAGGAAATCGGAAATTTGAGCCATGTGGCT 988  
DB 134 ATGACATTAAGTAATTCAGTATATCAAAAGTTATCCAGAAATTTGATTTATGAC 193  
QY 989 GCGAATTAACAAGGCGCTCAGCAGATATGATGACTGTGAGACTGTT 1038  
DB 194 GCGAATTAACAAGTACAGATATGTTTGTATATTCAGAAAGTGGCT 243

## RESULT 15

US-09-864-761-6494/c  
Sequence 6494, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: A60MGA-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 6494  
LENGTH: 491  
TYPE: DNA  
ORGANISM: Homo sapiens

FEATURE:  
OTHER INFORMATION: MAP TO AC004889.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3

